



INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI.

19426

R. I. S.
PC-84-10 AB-2K6-49-1,000.

PROCEEDINGS
OF THE
ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL 184

LONDON

Printed and published for the Royal Society
By the Cambridge University Press
Bentley House, N.W.1

6 November 1945

PRINTED IN GREAT BRITAIN BY
WALTER LEWIS, M.A.
AT THE CAMBRIDGE UNIVERSITY PRESS

CONTENTS

SERIES A VOLUME 184

No. A 996. 23 July 1945

	PAGE
Métabolisme de la thyroxine et de l'iode ionique. (Abstract.) By F. Joliot .	1
Questions of measurements and standardization in radioactivity. (Abstract.) By I. Joliot-Curie	2
Some new investigations on cosmic ray showers. (Abstract.) By P. Auger .	2
The infra-red spectra of polymers and related monomers. I. By H. W. Thompson and P. Torkington	3
The infra-red spectra of polymers and related monomers. II. By H. W. Thompson and P. Torkington	21
The topography of crystal faces. I. The topography of a (100) face of a left-handed quartz crystal. By S. Tolansky. (Plate 1)	41
The topography of crystal faces. II. The topography of cleavage faces of mica and selenite. By S. Tolansky. (Plate 2)	51
The crystal structure of cholesteryl iodide. By C. H. Carlisle and D. Crowfoot. (Plate 3)	64
The electrical conductivity of stannous sulphide. By J. S. Anderson and M. C. Morton	83
Reversible adsorption of proteins at the oil/water interface. I. Preferential ad- sorption of proteins at charged oil/water interfaces. By J. J. Elkes, A. C. Frazer, J. H. Schulman and H. C. Stewart. (Plates 4, 5)	102

No. A 997. 21 August 1945

Supersonic dispersion in gases. II. Air containing water vapour. By M. Mokhtar and E. G. Richardson	117
Complex potentials in two-dimensional elasticity. I. By A. C. Stevenson . .	129
Stress systems in anisotropic plates. III. By A. E. Green and G. I. Taylor, F.R.S.	181
Short period fluctuations in the characteristics of wireless echoes from the iono- sphere. By T. L. Eckersley, F.R.S. and F. T. Farmer. (Plates 6, 7) . . .	196
Complex potentials in two-dimensional elasticity. II. By A. C. Stevenson . .	218

No. A 998. 21 August 1945

	PAGE
Stress systems in isotropic and aeolotropic plates. V. By A. E. Green	231
Bakerian Lecture. On relaxation methods: A mathematics for engineering science. By R. V. Southwell, F.R.S.	253
Stress systems in aeolotropic plates. VI. By A. E. Green	289
Stress systems in aeolotropic plates. VII. By A. E. Green	301

No. A 999. 6 November 1945

The phosphorescence of various solids. By J. T. Randall and M. H. F. Wilkins. (Plate 8)	347
Phosphorescence and electron traps. I. The study of trap distributions. By J. T. Randall and M. H. F. Wilkins	365
Phosphorescence and electron traps. II. The interpretation of long-period phosphorescence. By J. T. Randall and M. H. F. Wilkins	390
Short period phosphorescence and electron traps. By G. F. J. Garlick and M. H. F. Wilkins	408
On the catalytic cyclization of aliphatic hydrocarbons. I. By E. F. G. Herington and E. K. Rideal, F.R.S.	434
On the catalytic cyclization of aliphatic hydrocarbons. II. By E. F. G. Herington and E. K. Rideal, F.R.S.	447
The age constitution and the future fellowship of the Royal Society. By L. Solomon	464
Index	479

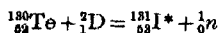
Métabolisme de la thyroxine et de l'iode ionique

By PROFESSOR F. JOLIOT

Read 22 February 1945

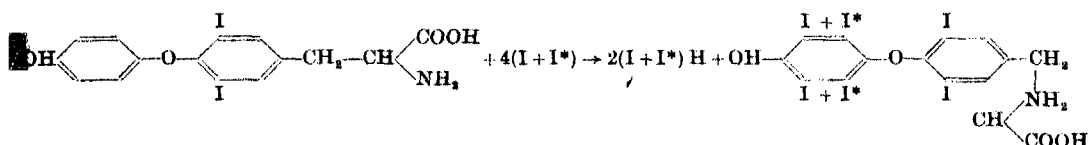
Sur l'obtention de la thyroxine contenant du radioiode et son comportement dans l'organisme.

(1) Préparation à l'aide des deutérons émis par le cyclotron du Collège de France de $^{131}_{53}\text{I}$ radioactif de période 8 jours. Irradiation du tellure par deutérons:



radioactivité $^{131}_{53}\text{I}^* \rightarrow ^{131}_{54}\text{Xe} + \text{électron négatif.}$

(2) Préparation thyroxine marquée par $^{131}_{53}\text{I}^*$ radioactif iode + diiodothyronine dissoute dans solution ammoniacale concentrée [diiodothyronine est l'avant dernier terme de la synthèse de la thyroxine par Barger & Harington (1927)]



0.1×10^{-6} g. de thyroxine est facilement décelable par la mesure des électrons de désintégration des atomes d'iode.

La détection de la thyroxine marquée par la radioactivité du radioiode qu'elle contient est beaucoup plus sensible et précise que par les méthodes antérieures (métamorphose des têtards, métabolisme de base, etc.).

(3) Injecte thyroxine en solution sodique dans le sang (veine marginale de l'oreille) de lapines gestantes.

(4) Sacrifie animaux 5 heures après injections et mesure activité dans: Sang, urine, bile, thyroïde, hypophyse, embryons, plasma du sang, globules du sang.

(5) Expérience analogue sur animaux dans les mêmes conditions physiologiques mais avec radioiodure de sodium.

(6) Dans les deux cas on dose chaque fois que cela est possible iode ionique, diiodothyrosine, thyroxine (technique de séparation Leland & Forster).

(7) Résultats principaux: Iode ionique se répartit également entre plasma et globules du sang. Thyroxine est presque en totalité dans le plasma.

La thyroxine est éliminée intensément par les reins beaucoup plus que l'iode ionique (à nombre d'atomes I radioactif injectés très voisin).

L'iode ionique se précipite avec un tropisme extraordinaire dans la thyroïde (ce qui était déjà connu) et 5 heures après l'animal excrète déjà de la thyroxine et de la diiodothyrosine synthétisées par l'organisme.

Il fut démontré autrefois que la thyroxine introduite dans l'organisme provoque la mise au repos de la glande thyroïde. On peut invoquer deux possibilités pour expliquer cette action: ou action directe de l'hormone sur la glande thyroïde, ou action directe de l'hormone sur la glande hypophysaire ou les deux. On a constaté que la thyroxine pénètre plus facilement dans l'hypophyse que ne le fait l'iode ionique, elle pénètre peu dans la thyroïde beaucoup moins facilement que l'iode ionique.

La méthode permet d'aborder le problème controversé de la traversée placentaire des hormones. Nos premiers résultats indiquent que, vers le milieu de la gestation (15 ou 16ème jour chez la lapine) l'iode ionique traverse le placenta plus facilement que la thyroxine et que celle-ci passe sans doute très lentement à cette époque de la gestation.

Ce travail a été fait en collaboration par: Messieurs R. Courrier, endocrinologiste; A. Horeau, chimiste organicien; P. Sue, chimiste inorganicien; F. Joliot, physicien.

Questions of measurements and standardization in radioactivity

By PROFESSOR IRÈNE JOLIOT-CURIE

Read 22 February 1945

For the physician, a radioelement is characterized by its period and the kind of rays it emits. What matters in most experiments is to know the quantity of a radioelement expressed in 'number of disintegration per second' rather than in weight. Generally, one expresses this quantity taking for unit the 'curie', which is the number of disintegrations per second of a gram of radium, and which has been carefully measured.

In some cases, the measure in curie can be directly obtained from α -ray measurements of a radioelement, but in general it is not possible to get it in that simple way.

Most of the absolute measurements depend on the International Radium Standard, made of pure salt of radium carefully weighed. The comparisons with the radium tube are made by the ionization of γ -rays. Special ionization or calorimetric measurement allows one to establish the correspondence between the 'milligram equivalent' measured in that way and the absolute value of a source, in millicuries.

For the artificial radioelements where no α -rays are available, the measurements are very unsatisfactory, and depend, in the best cases, on numeration of β -rays, which are difficult to make with precision.

Some new investigations on cosmic ray showers

By PROFESSOR P. AUGER

Read 22 February 1945

Some experiments performed by my pupils and myself during the last two years have yielded results which seem difficult to reconcile with the pure cascade theory of cosmic-ray showers. Among these are: the presence of penetrating particles (mesons) in extensive atmospheric showers (A-showers), the variation with altitude of the frequency of these A-showers, and the special structure of some local showers.

Using a meson selector in coincidence with distant counters, Dr Rogozinski could explain his results by: (1) knock-on electrons accompanying a meson; (2) mesons in A-showers; (3) multiple simultaneous mesons (meson shower). This last phenomenon being frequent only in high altitude.

The variation of the number of A-showers with altitude was studied by Dr Rogozinski and myself by the use of two sets of counters, which were sent to a height of 15 km., registering simultaneously the extensive and the ordinary local atmospheric showers. A comparison of the counting rates of the two sets at sea level and at high altitude suggested that the A-showers are generally born at less than 15 km. altitude. This would call for an explanation of the appearance of very high energy electrons or photons at these levels.

The structure of local showers connected with high energy cosmic-ray phenomena was studied by Dr Dandin with a cloud chamber controlled by a group of non-coplanar coincidence counters separated by thick lead screens. He found an anomalous frequency of narrow showers, and a small number of pair-showers, consisting of five or six pairs, of very small angular spread, rather regularly distributed in space. The interpretation of these effects seems to involve some kind of explosive process.

The infra-red spectra of polymers and related monomers. I

By H. W. THOMPSON, *Fellow of St John's College, Oxford*

AND P. TORKINGTON, *formerly Scholar of St John's College, Oxford*

(Communicated by C. N. Hinshelwood, F.R.S.—Received 10 March 1944)

A survey has been made of the infra-red absorption spectra of a number of polymeric substances and their associated monomers. The present paper deals with the hydrocarbon-type polymers such as polythene, polyisobutylene, buna, hydro-rubber, polystyrene, and compounds related to them. The results reveal the existence of methyl groups in polythene, and also of a small number of carbonyl groups. The type of unsaturated products formed during the cracking of polythene is also indicated. The variation of spectrum with chain length has been examined using samples of polythene and isobutylene. The influence of state of aggregation has been studied and the use of plane polarized radiation with oriented films of polythene has been explored.

INTRODUCTION

One of the most interesting features of polymeric substances, synthetic rubbers, and similar compounds of high molecular weight, is the wide variety of physical properties shown by the different groups of compounds. Even within the same class of polymer, small differences in the method of preparation, or proportions of the constituent parts, may result in marked physical differences, by which the value of the compounds may be determined. Physical properties must ultimately be closely connected with internal molecular structure, as well as by the factors which determine the cohesion between long chains or the packing of closed rings. If it were possible in a few cases to correlate more precisely such physical properties as elasticity, ductility, or electrical qualities, with particular characteristics of internal molecular structure, it might become possible to synthesize compounds having any particular set of physical properties.

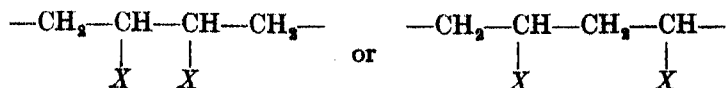
Very few such correlations have yet been possible, and further knowledge about the structure of polymers is needed. The nature of these compounds, however, often limits the application of the usual methods of physical chemistry, although the use of X-rays and certain other methods have given some striking results (Mark, Fuoss, Marvel & Ott 1943; and Melville 1941). Most methods used for the control and direction of large-scale production are still empirical. Measurement and analysis of the infra-red absorption spectrum might also have been regarded as an unpromising approach. Thus, substances of high molecular weight will have to be studied either as solids or in solution, and, in addition, their molecules will have very large moments of inertia. Both these circumstances preclude any analysis of rotational fine structure which may accompany the absorption of vibrational quanta. For practical purposes, therefore, only the vibrational spectra will be observed. In principle, each molecule will have a very large number of vibrational modes, and one might expect the spectrum to be so complicated that it could not be dissected.

Further, although a mathematical treatment of the vibrations of long-chain systems or lattice structures can be carried out by approximation methods, it has not hitherto proved precise enough to be useful for the purposes involved.

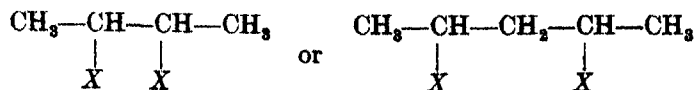
A study of the infra-red absorption spectra of series of structurally related molecules suggested that in spite of the above arguments, progress might be made by correlating the spectra of polymeric substances with those of simpler related molecules. It seems too that in many cases the vibrational spectra of polymers in the region $1-25\mu$ are simplified by two factors; first, the occurrence of recurring units may lead to a grouping of oscillation frequencies around a few particular values, and secondly, many vibration bands of the long-chain systems lie at lower frequencies outside this range.

It is now well known that characteristic vibration frequencies are associated with certain groups such as $-\text{CH}_3$, $>\text{CH}_2$, $\equiv\text{CH}$, $\text{O}-\text{H}$, $\text{C}=\text{O}$, $\text{C}=\text{C}$, $-\text{NH}_2$, $-\text{NO}_2$, and that these frequencies remain almost unchanged in magnitude in different compounds. Such groups can be detected and even estimated spectroscopically. More important, however, is the possibility that in a polymer, the skeleton of which is for the most part built up from a recurring unit, this unit shall have its own characteristic set of vibration frequencies which are not much affected when the units are joined together. It has been found, for example, that the tertiary butyl grouping $(\text{CH}_3)_3\text{C}-$ has characteristic absorption bands in the region of 8μ which are only slightly displaced by a change in the residue to which the group is attached. Again, with molecules of the type $\text{R}_1\text{CH}:\text{CR}_2\text{R}_3$, there is intense absorption in the infra-red due to the motion of the skeleton in the neighbourhood of the hydrogen atom, and this absorption varies characteristically with the nature of the radicals R_1 , R_2 , R_3 . This will be discussed below.

Again, when an unsaturated compound of the type $\text{CH}_2:\text{CH}\cdot\text{X}$, such as vinyl chloride, styrene, or butadiene, is polymerized, the resulting structure might be a head to head or head to tail arrangement, giving



In this case, comparison of the vibrational spectrum of the polymer with those of the simple molecules such as



may help to decide between the different alternatives, or to confirm evidence obtained by other methods. Further, if the polymerization does not proceed exclusively according to one pattern, it may be possible to determine from the observed absorption intensities the relative proportions of the different types of skeleton unit which are present, and this applies particularly to the case of interpolymers.

The infra-red spectra of various rubber polymers and related substances have been described by Stair & Coblenz (1935), by Sears (1941), and by Williams (1936). Wells (1940) has surveyed the spectra of various substances in a search for windows for the infra-red, and Barnes, Liddel & Williams (1943) have described measurements with synthetic rubbers particularly from the standpoint of analysis and plant control. No attempt to analyse and correlate the spectra systematically has yet been made. We have therefore begun a survey of the spectra of different classes of compound, and of the associated monomers and related simpler molecules. The measurements have enabled us to work out the best experimental methods, and have revealed the most promising lines for more detailed study. So far, most of our measurements have been confined to the region $5\text{--}20\mu$, which includes most vibration frequencies other than those determined by the stretching of C—H bonds. The resolving power which we have so far used is not great enough to reveal fine details in the region of 3μ . The conclusions so far reached are in many cases incomplete, but so many data have been accumulated that it seems desirable to summarize them now, so as to simplify later and more detailed accounts.

The substances which have been measured can be divided into the following classes: (1) hydrocarbon-type polymers such as polythene, polyisobutylene, rubber and polystyrene; (2) halogen-containing polymers such as polyvinyl chloride, polyvinylidene chloride, neoprene and halothenes; (3) oxygen-containing substances such as polymethyl acrylate and methacrylate, polyvinyl alcohol, polyvinyl acetate, and substances containing the peptide linkage such as nylon. While this rather arbitrary classification is not based on structural analogies, it is the most convenient for preliminary purposes. The present paper deals with the first class.

EXPERIMENTAL METHOD

The polymers were usually examined as thin solid films. In some cases they were supplied ready for use, and in others they were made by evaporation of solutions from the surface of mercury, or from a heated glass plate. The thickness of a film could be controlled by correct choice of concentration or amount of solution used. The particular solvents used for the different polymers are stated in the individual cases below. Waxy substances were used as thin solid layers between a pair of rock-salt plates, or as liquids in a heated cell made by placing nichrome heating coils around the edges of two rock-salt plates separated by a suitable washer. Since the heater gave a detectable infra-red emission, it was placed on the side of the cell away from the spectrometer.

Two spectrometers were used. One was a Hilger automatically recording Littrow spectrometer with 30° rock-salt prism, of about 100 cm. length of refracting face. A very sensitive Hilger-Schwarz differential vacuum thermocouple of 90 ohms resistance was used as detector, and the degree of resolution of water and ammonia absorption lines used in the calibration showed that the resolving power of this instrument was high. Deflexions of a moving-coil galvanometer of 10 ohms resist-

ance and sensitivity 150 mm./ μ A at 1 m., used as primary, were amplified by means of a twin photocell amplifier, which fed a Tinsley moving-coil galvanometer of 450 ohms resistance and sensitivity 1200 mm./ μ A at 1 m. The recording was taken from the latter photographically on a camera drum, geared to the mechanism which rotated the prism.

For the regions 14–20 μ and at wave-lengths below about 7 μ , a Hilger D 88 spectrometer was used, modified by the introduction of a Schwarz thermocouple and amplifier similar to the above. This spectrometer has interchangeable prisms of quartz, fluorite, rock salt, and sylvine, and although its resolving power with a rock-salt prism is not quite so high as that obtained with the above Littrow spectrometer, it has the advantage of the sylvine prism for the region 14–20 μ , and also the greater dispersion of fluorite for the region just below 7 μ . It is also more satisfactory than the recorder for use in the region around 6 μ , where there are absorption bands due to atmospheric water vapour.

THE SPECTRUM OF POLYTHENE

Figure 1 shows the absorption spectrum of a sample of polythene, the mean molecular weight of which, measured by viscosity, was stated to be 13,000. Two oriented films were used, 0.05 and 0.30 mm. in thickness. The absorption percentages shown in figure 1 cannot be taken as strictly accurate for two reasons. First, the films were not uniform in thickness over their whole area, and an average value was taken. Secondly, the irregularities in the surfaces of the films caused a marked general scattering, so that comparisons of the transmission through the film and through air were not quite valid. This scattering has the effect of raising the 'base-line' of the absorption by an indeterminable amount; in the curves shown an attempt has been made to correct the absorption curves actually measured so as to allow for this scattering error. The striking feature of the spectrum is the transparency of polythene over wide regions. This presumably arises, partly at least, from the essentially non-polar nature of the substance, since the intensity of absorption is connected with the variation of electric moment during vibration. Most other polymers show intense absorption over a wide spectral range, even with much thinner films.

The absorption at about 2900 cm.⁻¹ (3.4 μ) is due to absorption of frequencies associated with the vibrations of C—H bonds. It has been examined in detail by Fox & Martin (1940) for extracts of polythene in carbon tetrachloride, using the higher resolving power of a grating spectrometer. The dispersion and resolving power which we have so far used for this region does not bring out the fine details of this group of bands. Next in intensity are the two bands at 6.85 and 13.85 μ . The former is characteristic of all compounds containing CH₂ groups, such as paraffins and naphthenes, and is connected with a deformation vibration of this group. The band at 13.85 μ , which appears as a doublet with the solid films of polythene, is almost certainly connected with a vibration (or close pair of vibrations) of the long carbon-carbon chain. Normal paraffins with more than four carbon atoms in the chain, or

any other paraffins whose molecule contains a similar long unbranched chain, have a single strong band at this wave-length, and we have recently detected it also in other long-chain molecules such as stearic acid and dodecyl compounds. Figure 2 shows the absorption spectra of normal hexane, normal heptane, normal decane and a mixture of octadecane with nonadecane, over the relevant range of wave-lengths. With each of these substances the bands at 6.85 and 13.85μ are prominent, and with each of the liquid paraffins the latter band is single, and not a doublet as with solid polythene. This point is referred to again below. It is important to note that the detailed mathematical treatment of the vibrations of long-chain paraffins by Whitcomb, Nielsen & Thomas (1940) fails to account satisfactorily for the strong band at 13.85 , although this is one of the main features of their spectra. Since their calculations referred primarily to an infinite chain, the disagreement of their deduction with the established band of polythene is even more striking.

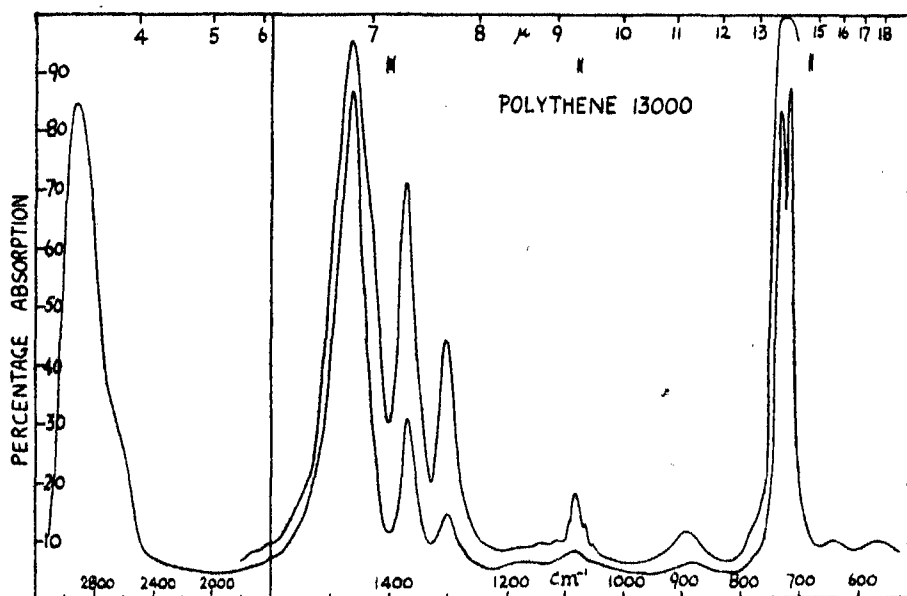


FIGURE 1

The other definite bands in the spectrum of polythene are at about 7.25 , 7.65 , 9.25 and 11.25μ , although there is some feeble absorption at other wave-lengths. The band at 9.25μ is probably composite; that at 7.65μ is probably connected with a rocking or bending motion of the methylene groups, like the band at 6.85μ .

The band at 7.25μ is particularly interesting, since it lies at the wave-length characteristic for absorption of the symmetrical deformation frequency of the methyl group. Comparison of the relative intensities of the absorption at 7.25 and 6.85μ in the case of a large number of normal paraffins, singly, doubly and multiply branched paraffins, and naphthenic hydrocarbons, has shown that the relative

intensity of the band at 7.25μ increases as the proportion of methyl groups in the molecule is increased, and this band is absent when no methyl groups are present. The absolute extinction coefficient per methyl group is higher than that per CH_2 group, so that even when there are many more CH_2 groups than CH_3 groups, the two bands have roughly equal intensity. In the sample of polythene used for figure 1, therefore, there are relatively few, but certainly some, methyl groups. This provides an independent confirmation of the earlier conclusion of Fox & Martin, which was obtained from a study of the absorption bands at about 3.4μ .

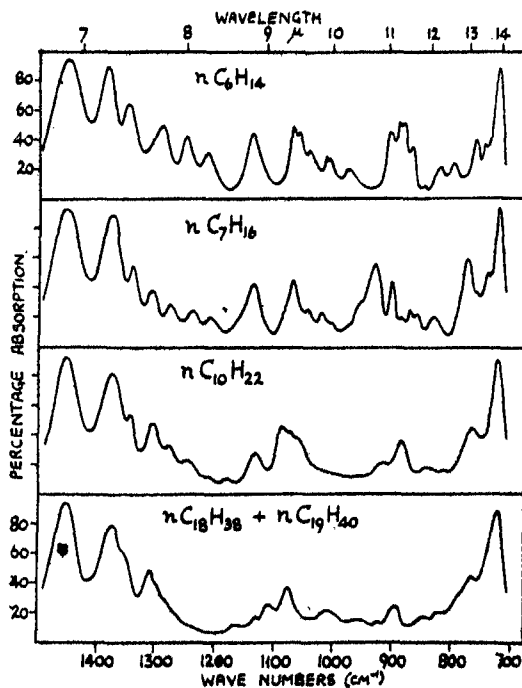


FIGURE 2

POLYTHENES OF DIFFERENT MOLECULAR WEIGHT AND CRACKED POLYTHENES

With the lower normal paraffins, there are definite alterations in the absorption spectrum on passing from one member to the next higher homologue. With more than about 8 carbon atoms the spectra become progressively more alike. It seemed desirable, however, to see whether any systematic variation could be detected in the spectra of polythenes of different chain length. For this purpose, two sets of samples were obtained. The first included: (i) a xylene extract, ref. no. 25/B/534, molecular weight 7100, supplied as a solid film; (ii) a waxy ether extract, molecular weight 2400; and (iii) a wax, molecular weight 1400, ref. 'Catalytic Process'. The second set of four specimens was a group obtained by cracking polythene of high

molecular weight, three being waxes of molecular weights 500, 700 and 1100, and the other a solid film of molecular weight 5000.

The absorption spectra of the first set over the range 5–14 μ are shown in figure 3. Exactly equal thicknesses could not be obtained, and this, together with the irregular scattering already mentioned, makes it difficult to compare accurately the absorption percentages of the main bands. The curves shown have been obtained by recalculation from the measurements so as to correspond to a constant thickness of about 0.1 mm. Although the main bands of high molecular weight polythene at

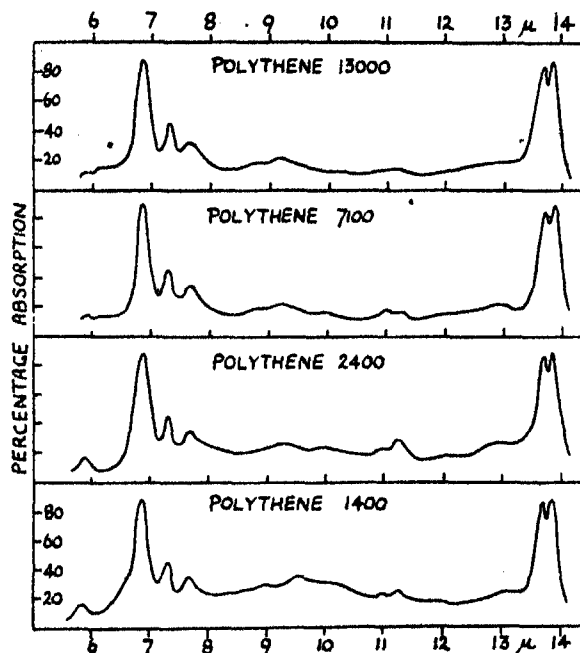


FIGURE 3

6.85, 7.25, 7.65 and 13.85 μ are reproduced in the spectra of figure 3, there are definite differences among the feebler bands with the different samples. These variations might indeed arise because the different lengths of chain have different vibrational spectra, but there are no obvious regularities, and it will become clear from the other results given below that the results are better explained if the different samples contain varying amounts of key groups.

The spectra of the cracked polythenes are shown in figure 4. They form a regular series among themselves, and also show noticeable differences from the samples of figure 3. Here again, owing to the unequal thicknesses of the films used, the curves have been corrected so as to be roughly comparable. It is seen that the cracked polythenes show the same key bands at polythene 13,000, namely, at about 6.85, 7.25, 7.65 and 13.85 μ . In other regions, however, there are marked differences from the parent polythene, particularly in the region of 8 μ , and between 10 and 11.5 μ .

As the molecular weight decreases, a band at 11.02μ , not detectable in polythene 13,000, increases steadily in intensity, and the same applies to a band at 10.1μ . Another band at 10.35μ is found with the cracked polythenes, but is probably absent from the spectrum of polythene 13,000. The band at 11.25μ , well marked with the cracked samples, might correspond to that found with polythene 13,000, but this is unlikely. Also, the broad maximum found with polythene 13,000 at about 9.25μ may be displaced towards longer wave-lengths as the molecular weight is decreased. In this respect there was a difference between polythene 13,000 and the sample of molecular weight 1400, ref. no. 'Catalytic Process'.

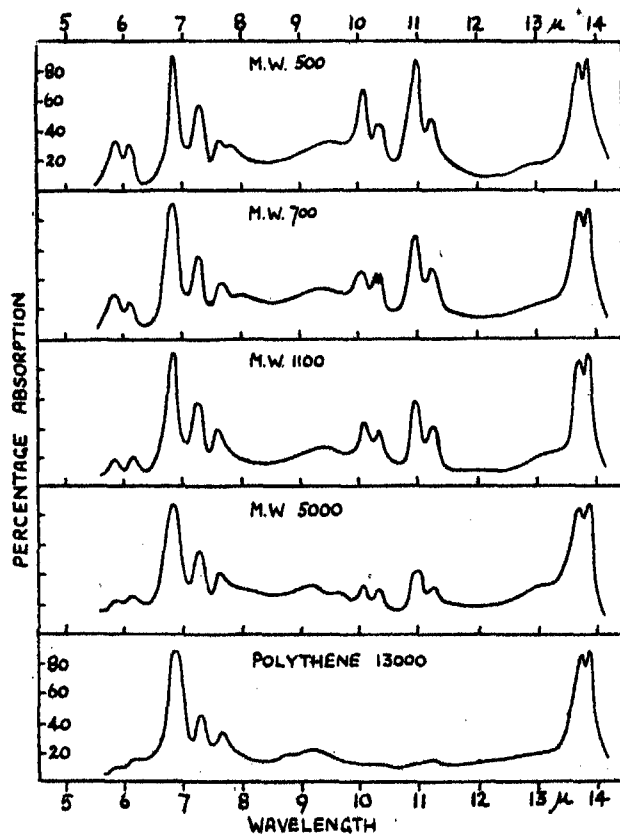
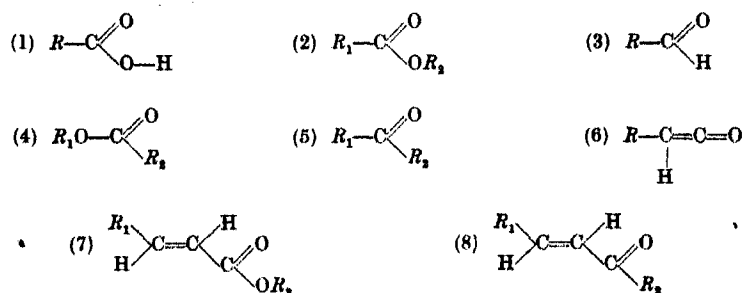


FIGURE 4

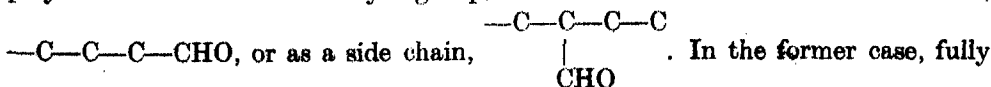
Two other bands, at 5.85 and 6.1μ , are well marked in the spectra of the cracked polythenes. These peaks, at 1710 and 1640 cm^{-1} respectively, are too widely separated to be branches of the same band. Moreover, their relative intensity varies in the different samples, suggesting that they are separate vibration bands. The lower frequency 1640 corresponds to what would be expected for the valency vibration of a normal carbon-carbon double bond, and that at 1710 to the corresponding vibra-

tion of a carbonyl linkage. This interpretation is supported by the facts that (i) the intensity of absorption due to the carbon-carbon double bond increases more or less uniformly as the cracking proceeds, more unsaturated materials being formed, (ii) the spectrum of oxidized polythene given below shows the peak at 1710 with high intensity.

The vibration frequency associated with the C=C bond is not much affected by the attached groups, unless there is a conjugated system such as —C=C—C=O . The normal value of the frequency actually observed in the present case therefore excludes such a conjugated structure with the carbonyl group. On the other hand, the vibration frequency of the carbonyl group varies appreciably with the radicals attached to the carbon atom, as Kohlrausch has shown. Possible structures which might be considered for the cracked polythenes are the following:



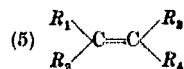
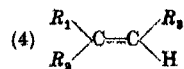
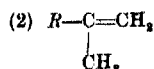
Most of these can be excluded on general grounds, or for spectroscopic reasons, or both. Thus, the observed carbonyl group absorption band is at 1710 cm^{-1} . The Raman frequency of the carbonyl link vibration in carboxylic acids is about 1650, and conjugation with an unsaturated bond as in (7) or (8) also leads to low values. The only skeleton structures which fit the facts are (3) and (5). Aldehydes have a Raman interval of about 1720, and ketones one of about 1710. It is impossible to decide at present between the two alternatives. If, however, the polythenes contained an aldehyde group, it could occur either at the end of a chain,



oxidized polythene might be expected to have some carboxylic groups. This is not suggested by the spectrum. If side-chain aldehyde groups were present, these might be formed during the oxidation of polythene by oxidation of side methyl groups, leading to a reduction of the absorption by these groups. This, again, does not seem to occur, and the balance of evidence suggests that the cracked polythenes may contain the ketonic chain —C—CO—C— .

The cracked polythenes have four absorption bands between 10 and 11.5μ , namely, at $990, 966, 909$ and 889 cm^{-1} . All four bands increase in intensity as the cracking proceeds, and might therefore be expected to be connected with vibrations of an unsaturated skeleton; and, further, by vibrations which are determined by

nuclei close to the carbon-carbon double bond. Several nuclear skeletons can be visualized for the region of the C=C bond. These are:



With polythenes, the tetra-alkyl olefine type (5) is so improbable that it need not be considered further. The infra-red absorption spectra of hydrocarbons of each of the other types have recently been measured, and can be compared with those of the cracked polythenes. Table 1 lists some of the compounds so far examined; substances like isoprene and piperylene are allocated on the basis of their molecular structure to two classes of nuclear skeleton.

TABLE 1

Class (1)	Class (2)
propylene	isobutylene
butene-1	methyl ethyl ethylene
pentene-1	2.5-dimethyl-1.5-hexadiene
octene-1	2.2.4-trimethyl-pentene-4
isopropyl ethylene	isoprene
styrene	1.3-dimethyl-1.3-butadiene
1.3-butadiene	
1.4-pentadiene	
isoprene	
1.1-dimethyl-1.3-butadiene	
piperylene	
Class (3)	Class (4)
butene-2	trimethyl ethylene
pentene-2	2.4.4-trimethyl-pentene-2
octene-2	2.5-dimethyl-2.4-hexadiene
piperylene	1.1-dimethyl-1.3-butadiene
1.3-dimethyl-1.3-butadiene	

In figure 5 the spectra of examples of each class are shown over the range 9-14 μ . Inspection of the curves of figure 5 reveals the striking fact that for all members of class (1) the two most intense bands in the spectrum lie close to 910 and 990 cm^{-1} , so that these bands may be used as a characteristic of the vinyl radical $-CH=CH_2$ in hydrocarbons. Similarly, compounds of class (3) have an intense characteristic absorption between 960 and 970 cm^{-1} , and those of class (2) have one at about 890 cm^{-1} . Detailed considerations with some of the simpler olefines, which will be described elsewhere, shows that these characteristic vibrations are associated with deformational motions of the CH bonds in the skeleton. Similar results have been found with molecules containing a similar skeleton, such as vinyl chloride, vinyl bromide, vinylidene chloride, cetene and stearic acid, and these will be given in

other papers of this series. With compounds of the class (4), there is no correspondingly well-defined vibration band, although few substances of this type have yet been measured.

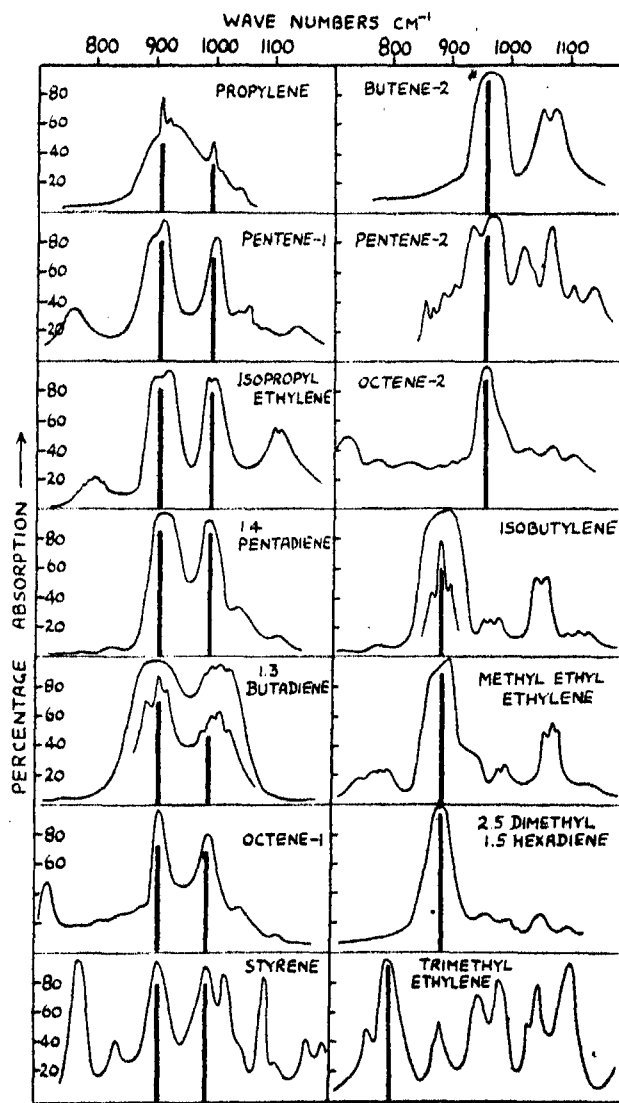


FIGURE 5

This differentiation between the different types of unsaturated structure is also borne out by the Raman spectra of the hydrocarbons concerned. It must be remembered that in principle the Raman frequencies of a molecule may not be directly comparable with those deduced from the infra-red spectrum, since different selection rules come into play, and frequencies which are strong in one case may be weak in

the other. It is found, however (Kohlrausch 1938; Bourguet & Piaux 1935; Piaux 1935; Gredy 1935), that olefines of class (1), such as butene-1, pentene-1, nonene-1 and allyl cyclopentane, all show frequencies of about 910 and 990 cm^{-1} , and those of class (3), such as pentene-2, hexene-2, heptene-2 and octene-2, show one of about 970. The Raman data suggest that the latter frequency is probably connected with the *cis* form, and further infra-red measurements on the *cis* and *trans* forms may lead to a more exact correlation than that given here. Similarly, 2,3-dimethyl-1,3-butadiene, of class (2), has a Raman frequency of 893. Dr R. Rasmussen has recently informed us that he has also reached similar conclusions to the above from measurements on the infra-red spectra of a large number of olefines.

Now the strong bands which appear in the cracking of polythene are at about 990, 966, 909 and 889 cm^{-1} . The intense pair at 990 and 909 can at once be correlated with the structure $R-\text{CH}=\text{CH}_2$, 966 with the structure $R_1-\text{CH}=\text{CH}-R_2$, and 889 with $R_1 R_2 \text{C}=\text{CH}_2$. It should be possible, if the necessary calibrations were first made, to estimate from the relative intensities of the different bands the proportions of the different structures which are formed in the cracking operation. It may be noted that the production during cracking of the structures of type (2), namely, $R(\text{CH}_3)\text{C}=\text{CH}_2$, is satisfactorily explained without migration of methyl groups if the original polythene sample contained methyl groups on the side of the main carbon chain.

We can now reconsider the spectra of polythene 13,000 and the waxes of lower molecular weight (figure 3). The spectrum of thicker films of polythene 13,000 (figure 1) shows very feeble maxima at 5.85 and 6.1μ corresponding to $\text{C}=\text{C}$ and $\text{C}=\text{O}$ linkages, although at very low concentration. The absolute extinction coefficient of the carbonyl group is known to be much higher than that of the $\text{C}=\text{C}$ linkage, and it can therefore be inferred from the rough equality of absorption of polythene at 5.85 and 6.1μ that this polymer probably contains fewer carbonyl groups than $\text{C}=\text{C}$ groups. Each of the waxes shows a feeble band at 5.85μ , which is more marked with the polymers 2400 and 1400. These bands are attributed to the carbonyl group, and it may be that the shorter chains are richer in this linkage. Most of the other feeble bands can be correlated with small amounts of unsaturated structures discussed above. Absorption at 5.85μ due to the $\text{C}=\text{C}$ bond may be masked in samples 2400 and 1400 by the neighbouring carbonyl group absorption.

THE SPECTRUM OF OXIDIZED POLYTHENE

The absorption spectrum of a thick solid film of highly oxidized polythene has been measured, with the following main results: (i) the carbonyl group absorption at 5.85μ is very intense; (ii) the band at 7.25μ characteristic of methyl groups is at least as pronounced as in ordinary polythene; (iii) a new band appears at about 8.5μ ; this is not found with any other polythene or cracked polythene, and may be connected with a vibration of the carboxyl group; (iv) the bands associated with an unsaturated skeleton are weak.

THE EFFECT OF STATE OF AGGREGATION

Since, as described above, comparisons had to be made between the spectra of liquid and gaseous paraffins and that of solid polythene, it was desirable to know how the spectrum of polythene is affected by melting. Four samples, of molecular weights 700, 1100, 2400 and 7100, were examined in the solid and liquid state. The spectrum is essentially unchanged by the change of phase, and any displacement of the wave-lengths of the main bands must be very small. It is interesting, however, that the doublet bands at 13.85μ found with solid polythene always appear as a broader and shallower single band when the liquid is used. While no explanation can at present be given for this effect, it may be noted that the absorption bands of simple paraffins sharpen when the substances are cooled to very low temperatures and measured in the solid state (Avery & Ellis 1942).

THE USE OF POLARIZED INFRA-RED RADIATION

When long-chain polymers are available as oriented solid films, the use of plane polarized radiation for measuring the spectrum should throw some light on the molecular structure; for in these circumstances the relative intensity of different bands may change according as the plane of polarization of the incident beam is parallel or perpendicular to the molecular chains, and some information about the orientation of groups attached to the main chain may be obtained.

Plane polarized radiation was made by reflexion from a plane selenium mirror at the correct angle of incidence (about 70°). The degree of polarization for sodium D lines was measured and found to be about 97 %. According to Pfund (1906) this degree of polarization extends far into the infra-red. The loss of intensity at the selenium mirror was high, but using slightly wider slits the effect could be studied on the two strongest bands of oriented polythene at 6.85 and 13.85μ . The relative intensity of the two bands was measured with the direction of the oriented chains in the molecule successively in positions parallel and perpendicular to the plane of polarization of the incident radiation.

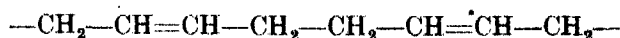
A small difference was found, but owing to the errors arising from general scattering by the film, its reality was doubtful. It is unfortunate that so few intense bands occur in the spectrum of polythene upon which the effect can be tested, and it is hoped shortly to use oriented films of other polymers containing more polar groups.

THE SPECTRA OF POLYISOBUTYLENE, BUNA AND HYDRO-RUBBER

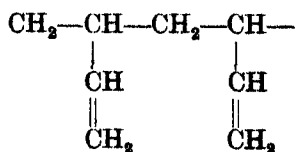
These spectra were measured not only for their intrinsic interest, but also in connexion with the determination of methyl groups in polythene, since the content of methyl groups in polyisobutene and hydro-rubber is known and might be used as reference standards. The spectra of isobutylene, 1.3-butadiene, isoprene and piperylene, will be described elsewhere soon.

A range of polyisobutenes was examined; these varied from viscous liquids of molecular weight about 9000 to rubbery solids of molecular weight about 100,000. No variations could be detected in the spectrum with changing molecular weight. The curves shown in figure 6 were obtained with films about 0.17 and 0.02 mm. thick, made by evaporation from a solution in petroleum ether from the surface of mercury. Polyisobutene is transparent over wide spectral ranges, such as 12–20 and 5–7 μ . The characteristic bands due to vibrations of methyl and methylene groups at about 7.25 and 6.8 μ are intense, but the methyl group vibration appears as a doublet, with components at 1370 and 1395 cm^{-1} . This doubling is not at present explained, but it may be noted that it occurs in a similar way with tertiary butyl benzene and isopropyl benzene but not with ethyl benzene or normal propyl benzene. This suggests that it occurs when two methyl groups are attached to the same carbon atom, and it is, in fact, found, though less markedly, in the spectra of 2.2-dimethyl-butane, 2.2-dimethyl-pentane, 2.2-dimethyl-hexane, 2.2.4-trimethyl-pentane and 2.2.3-trimethyl-pentane. The intense absorption of polyisobutene at about 1200 cm^{-1} is also similar to that of molecules containing the tertiary butyl grouping, or 3.3-dimethyl-pentane and 3.3-dimethyl-hexane; and both the bands at about 1200 and the doublet at about 940 cm^{-1} are almost certainly due to vibrations of the distorted C_4 tetrahedron in the molecular skeleton.

The spectrum of a sample of buna 85 is also shown in figure 6. This specimen was made by the polymerization of 1.3-butadiene in the presence of sodium. Solid films about 0.12, 0.08 and 0.02 mm. were prepared by evaporation from solutions in ethylene dichloride on the surface of mercury. The polymerization might proceed by 1.4 addition, giving



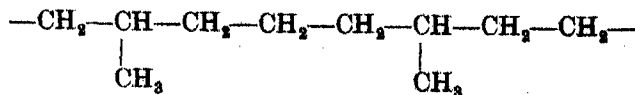
or by 1.2 addition, giving—



or a combination of both. The spectrum measured shows the $\text{C}=\text{C}$ oscillation at about 1660 cm^{-1} , and a variety of CH deformational modes between 1300 and 1500. It is possible that the $\text{C}=\text{C}$ vibration is double.

The intense bands at 905, 995 and 965 are more interesting. As described above, these indicate the presence of structures $R-\text{CH}=\text{CH}_2$ and $R_1-\text{CH}=\text{CH}-R_2$. Hence the polymerization has proceeded simultaneously by both 1.4 and 1.2 addition, and calibration of the band intensities would make it possible to determine the extent of each type.

Figure 6 shows the spectrum of hydro-rubber, which contains the skeleton



This substance is a highly viscous liquid, which was melted on to a rock-salt plate as a thin film and allowed to set. Some features of the spectra of branched chain paraffins such as isohexanes and iso-octanes are reproduced in this spectrum, but discussion must be postponed until the facts are more complete. For present purposes, it should be noticed that the absorption by methyl groups at 7.25μ is much stronger than in polythene, and it is not a doublet, as with polyisobutylene. The strong band at about 735 cm.^{-1} is displaced from the position at 725 cm.^{-1} found with straight-chain paraffins; this agrees with what is found for some singly branched paraffins such as 2-methyl-pentane.

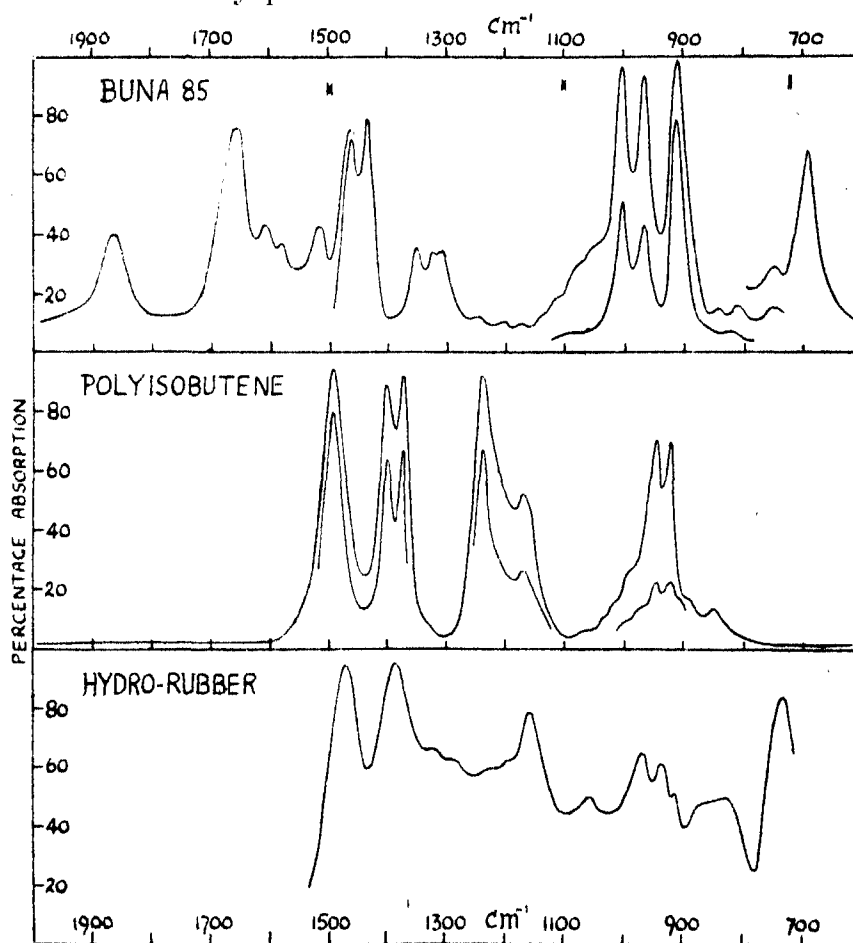


FIGURE 6

THE SPECTRUM OF STYRENE AND POLYSTYRENE

Figure 7 shows the spectrum of polystyrene between 5 and 20μ . The sample was made by the slow polymerization of styrene, and the solid films used for these measurements were 0.03 and 0.09 mm. thick, made by evaporation on mercury of

solutions in xylene. The spectrum is very complicated, and any detailed allocation of the vibration frequencies to particular modes would be purely speculative. We can, however, correlate the band at 758 cm^{-1} with a characteristic vibration of the phenyl group, and that at 1460 with the deformations of the CH_2 group. The spectrum shows certain similarities with those of isopropyl and tertiary butyl benzene, and noticeably fewer with that of normal propyl benzene.

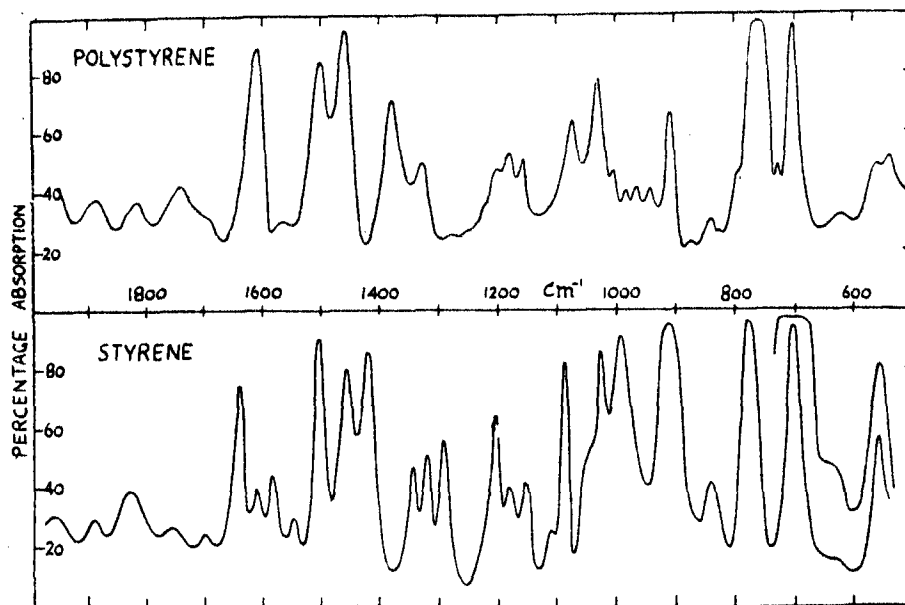


FIGURE 7

The Raman spectrum has been measured by Signer & Weiler (1932), and in table 2 the two sets of data are compared.

The band found at 1380 corresponds to that characteristic of the methyl group deformation, and might suggest the presence of CH_3 groups in polystyrene. This would imply the presence of the group $-\text{C}(\text{CH}_3)(\text{C}_6\text{H}_5)-$ in the chain. Although this must be regarded as probable, it must not be forgotten that the band at 1380 cm^{-1} might be explained as an overtone or combination frequency. Thus, there is an intense band at 701 , the overtone of which might be responsible for the band at 1380 .

Figure 7 also shows the spectrum of styrene. The marked differences would provide a means for following the rate of polymerization.

METHYL GROUPS IN POLYTHENE

The existence of methyl groups in polythene, revealed by the band at 7.25μ , is a matter of some importance. The ratio of CH_2 groups to CH_3 groups could be determined in either of two ways. If, as a first approximation, it is assumed that the optical extinction of groups such as CH_2 and CH_3 is independent of the rest of the

TABLE 2

(s. strong, m. medium, w. weak, v.w. very weak)

Raman	infra-red	Raman	infra-red
196 (2)	—	1094 (0, <i>d</i>)	1094 v.w.
211 (2)	—	1110 (2)	1115 v.w.
	{ 545 m.		1158 m.
	{ 560 m.	1177 (1)	1180 m.
620 (2)	623 w.	1204 (2)	1200 w.
	701 s.	1269 (0)	—
722 (0)	726 w.	1326 (0, <i>d</i>)	1328 m.
750 (0)	758 s.		1380 m.
	795 w.	1447 (1)	—
831 (0)	829 w.	1470 (0)	1460 s.
	842 m.		1500 s.
	876 w.	1588 (1)	1575 s.
	905 s.	1608 (3)	1610 s.
	941 m.		1695 v.w.
	967 m.		1745 w.
	984 w.		1820 w.
998 (4)	1005 w.		1895 w.
1032 (2)	1032 s.		1965 w.
1058 (0)		2936 (2, <i>d</i>)	
	1072 m.	3056 (2)	
		3066 (1)	

molecule, the absolute extinction coefficients could be determined using known thicknesses of simple paraffinic hydrocarbons of known structure. Using a known thickness of polythene the number of each type of group would then be measurable. The experimental difficulties involved in doing this satisfactorily have not yet been completely overcome, since it requires accurate measurement of very thin layers of liquids in a rock-salt cell; and there also remains the difficulty, referred to above, of irregular scattering by the polythene surface. Also, the film may not be uniform in thickness. It is also probable that in measuring the extinction coefficient at the two wave-lengths 7.25 and 6.85μ slight overlapping of bands will have to be taken into account. Attempts to carry out measurements of this kind are now in progress. Preliminary measurements using samples of polythene made in the presence of propylene showed a uniform change in the methyl group content as the propylene is taken up, and suggest that this method of analysis may be successful.

Meanwhile, a rough estimate of the methyl groups can be made by comparing the relative intensity of the bands due to CH_3 and CH_2 at 7.25 and 6.85μ in comparable thicknesses of lower paraffins, decane, octadecane, cetene, hydro-rubber, etc. This suggests that the polythene of high molecular weight used above may contain one methyl group for about fifty CH_2 groups.

The above work could not have been carried out without the co-operation of several firms in supplying the samples of polymers. We are particularly indebted to I.C.I. (Alkali) Ltd. for the polythenes and related substances, to I.C.I. (Dyestuffs)

Ltd. for a sample of buna 85, to the Anglo-Iranian Oil Co. Ltd. for several unsaturated hydrocarbons and a sample of polyisobutene, and to the Anglo-American Oil Co. Ltd. for a series of polyisobutenes.

We are also grateful to the Government Grant Committee of the Royal Society and to the Chemical Society for assistance in the purchase of equipment.

REFERENCES

- Avery & Ellis 1942 *J. Chem. Phys.* **10**, 10.
Barnes, Liddel & Williams 1943 *J. Industr. Engrng Chem.* **15**, 83.
Bourguel & Piaux 1935 *Bull. Soc. chim.* **2**, 1958.
Fox & Martin 1940 *Proc. Roy. Soc. A*, **175**, 208.
Gredy 1935 *Bull. Soc. chim.* **1**, 1030; **2**, 1951.
Kohlrausch 1938 *Der Smekal Raman Effekt*. *Ergänzungsband*.
Mark, Fuoss, Marvel & Ott 1943 *The chemistry of large molecules*. New York: Interscience Publ.
Melville 1941 *J. Chem. Soc.* p. 414.
Pfund 1906 *Johns Hopk. Circ.* **4**, 13.
Piaux 1935 *Ann. Chim.* (11), **4**, 147.
Sears 1941 *J. Appl. Phys.* **12**, 35.
Signer & Weiler 1932 *Helv. chim. Acta*, **15**, 649.
Stair & Coblenz 1935 *Bur. Stand. J. Res.* **15**, 295.
Wells 1940 *J. Appl. Phys.* **11**, 137.
Whitecomb, Nielsen & Thomas 1940 *J. Chem. Phys.* **8**, 143.
Williams 1936 *Physics*, **7**, 399.

The infra-red spectra of polymers and related monomers. II

BY H. W. THOMPSON, *Fellow of St John's College, Oxford*
AND P. TORRINGTON, *formerly Scholar of St John's College, Oxford*

(Communicated by C. N. Hinshelwood, F.R.S.—Received 10 March 1944)

The infra-red spectra of some polymers containing chlorine have been measured, and also those of the related monomers. Vibrational analyses and a study of band contours with vinyl chloride, vinyl bromide, vinylidene chloride, and vinylidene bromide, have made it possible to assign magnitudes to many of the normal vibration frequencies. The vibration frequencies of the series of vinyl halides have been correlated. The rotational structure associated with some of the vibration bands has been discussed in relation to the molecular structure. Similar measurements and considerations have been applied to chloroprene. The spectra of polyvinyl chloride, polyvinylidene chloride, and some halothenes and neoprenes have been examined. As a preliminary to the correlation of the absorption frequencies with vibrations of the nuclear skeletons, the spectra of some simple chlorinated paraffins have been measured.

In the previous paper the vibrational spectra of some hydrocarbon-type polymers were discussed. Similar measurements are described here on the parent monomers of some halogen-containing polymers, and on some of the polymers themselves. Valuable information about the normal vibration frequencies of the simple monomeric substances has been obtained, but there are still not enough data to justify a detailed consideration of the spectra of the polymers in relation to their molecular structure. In spite of this, the results are useful in many connexions, and serve to illustrate the stage reached in the general problem of the correlation of vibrational spectra and molecular structure.

VINYLDENE CHLORIDE

This substance, CH_2CCl_2 , is the monomer of polyvinylidene chloride, known commercially as saran. Apart from this, its spectrum has intrinsic interest, for, although the molecule is an asymmetrical rotator with three differing moments of inertia, it has one two-fold axis of rotation and two planes of symmetry, so that the infra-red absorption bands of the vapour should show an interesting contour. The moments of inertia are fairly large, but still small enough to make this rotational contour resolvable. Moreover, as the results show, it is possible here unambiguously to assign values to most, and perhaps all, of the normal vibrational modes, which is valuable in the general consideration of the vibration frequencies of vinyl and vinylidene compounds. A few infra-red bands have been measured previously by Emschwiller & Lecomte (1937), but their results are incomplete and the resolving power used by them was low.

The sample of vinylidene chloride was supplied by the research department of I.C.I. (Plastics) Ltd. The liquid was stabilized with a trace of thymol, and redistilled just before use, with a boiling-point of 31°C . The infra-red measurements were carried

out with the vapour, which polymerized so slowly as to cause no difficulties. In measuring the Raman spectrum, however, the liquid had to be used. The redistilled pure liquid polymerized rapidly, and produced a scattering of incident radiation which made accurate measurement of the Raman displacements impossible. It was finally possible to obtain stability by adding a trace of hydroquinone, too small in quantity to give detectable Raman emission. A Hilger E 518 spectrograph was used, with the mercury line at 4358 Å as incident radiation. Kodak Super Panchro Press plates were used.

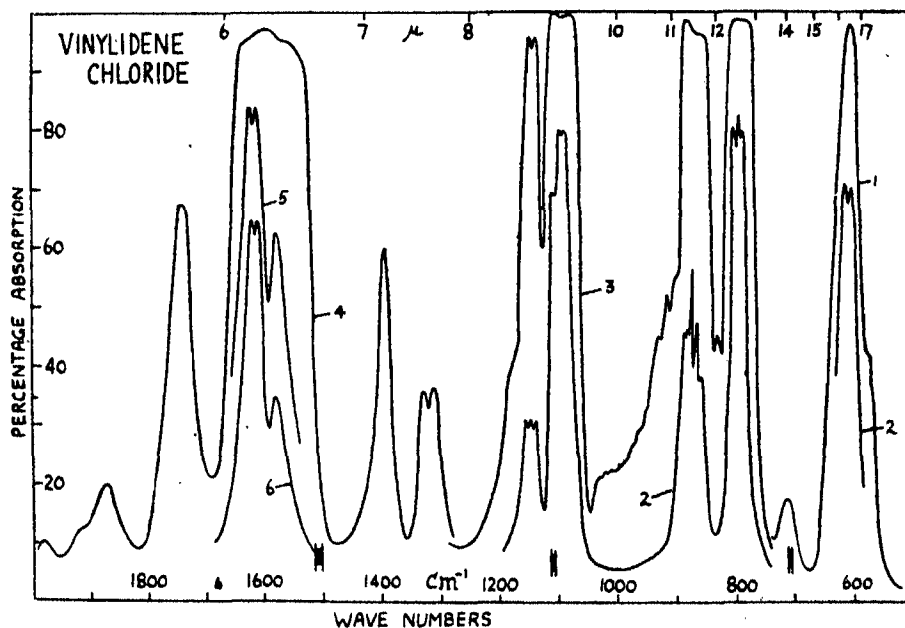


FIGURE 1. Vinylidene chloride. (1) 7 cm. path, 30 mm.; (2) 7 cm. path, 25 mm.; (3) 7 cm. path, 300 mm.; (4) 7 cm. path, 450 mm.; (5) 20 cm. path, 25 mm.; (6) 20 cm. path, 5 mm.

The infra-red spectrum between 5 and 20 μ is shown in figure 1 for various pressures of vapour in absorption cells 7 and 21 cm. in length. In order to obtain stronger absorption in the region of 3 μ , a film of liquid about 1 mm. thick was also used. The position of the absorption bands in wave numbers are given in table 1.

The curves of figure 1 show that many of the absorption bands have a well-marked rotational contour, the significance of which is discussed below. The following Raman frequencies were deduced from measurements on four plates: 295 (s.), 375 (m.), 450 (w.), 540 (w.), 595 (s.), 868 (v.w.), 1099 (?), 1331 (?), 1390 (m.), 1558 (m.), 1611 (s.), 3035 (m.) and 3130 (m.). Cabannes (1938) has recently given Raman data for this molecule. Our present results confirm his values of 293, 372, 600, 860, 1393, 1611 and 3038, but we have been unable to find displacements corresponding to three other frequencies given by him, namely, 146, 990, 1200 (?). Cabannes experienced

much difficulty due to polymerization, and from a consideration of our own Raman and infra-red data and their interpretation given below, these three frequencies must be regarded as spurious.

TABLE 1

wave number	intensity	wave number	intensity	wave number	intensity
580	very weak	904	weak	1391	medium
601	strong	914		1580	medium
612		931		1616	strong
714	very weak	970	very weak	1620	
787	strong	1011	very weak	1742	medium
794		1020		1876	weak
801		1030		1920	very weak
829	very weak	1089	strong	1988	very weak
858	strong	1094		2070	very weak
865		1103		2215	very weak
872		1136	medium	2430	very weak
877		1142		2670	very weak
884		1149		3035 (liq.)	weak
		1311	weak	3130 (liq.)	weak
		1326			

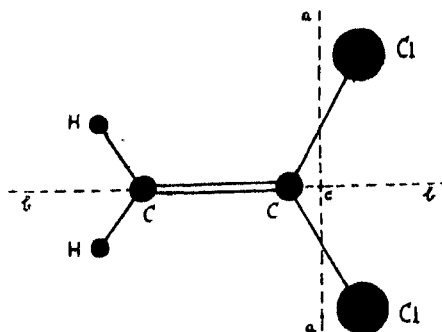


FIGURE 2. Vinylidene chloride.

Vinylidene chloride (figure 2) is planar (Brockway, Beach & Pauling 1935), with $r_{CH} = 1.08$ A, $r_{CC} = 1.38$ A, $r_{CCl} = 1.69$ A, and $\text{Cl}-\widehat{\text{C}}-\text{Cl} = 116^\circ$, $\text{Cl}-\widehat{\text{C}}=\text{C} = 122^\circ$, and $\text{H}-\widehat{\text{C}}-\text{H} = 120^\circ$. The moments of inertia thus calculated are about 105, 255 and 360×10^{-40} g.cm.². The least axis of inertia, *aa*, is in the plane of the molecule and perpendicular to the C—C bond; the middle axis, *bb*, is along the C=C bond, and the major axis, *cc*, is perpendicular to the plane of the molecule.

The molecule falls in the symmetry point group C_{2v} and there are twelve normal modes, which can be allocated to different symmetry classes as in table 2. In the latter, C_2 is the twofold axis of symmetry lying along the C=C bond, σ_v is the plane of the molecule, σ_v' the plane at right angles to the plane of the molecule and bisecting the HCH and ClCCl angles.

TABLE 2

sym- metry class	symmetry with respect to			permitted in Raman spectrum	permitted in infra- red	axis to which change of moment is parallel	type of infra-red band	number of modes	nature of modes
	C_2^x	σ_x	σ_y						
A_1	s	s	s	yes (p)	yes	middle	B	5	CH stretching C—Cl stretching C=C stretching CH ₂ deformation CCl ₂ deformation
A_2	s	as	as	yes (dp)	no	—	—	1	twisting
B_1	as	as	s	yes (dp)	yes	major	C	2	two non-planar deformations
B_2	as	s	as	yes (dp)	yes	least	A	4	CH stretching C—Cl stretching two planar rocking motions

Badger & Zumwalt (1938) have discussed the contour of vibration bands of asymmetrical rotators and differentiate A , B and C bands according as the change of electric moment is parallel to the least, middle, or greatest axis of inertia. In their nomenclature, the above values for the moments of inertia lead to $\sigma = 1.7$ and $S = -0.66$. Although a molecule of these particular dimensions was not considered by Badger & Zumwalt, extrapolation from their curves suggests that the A -type bands will have three submaxima, the spacing between the two outermost being about 14 cm.^{-1} . The B -type bands will consist essentially of a double band with central gap, the spacing between the main pair of submaxima being about $11\text{--}12 \text{ cm.}^{-1}$. The C -type bands will have a fairly strong central Q branch, with two other submaxima on each side of it (O , P , R and S). The spacing between the pair of outermost maxima, O – S , will be about 24 cm.^{-1} , and that between the next inner pair about 12 cm.^{-1} .

Now consider an allocation of the observed frequencies to the various normal modes. It has not yet been possible to measure the degree of depolarization of the Raman lines, but the totally symmetric vibrations may be expected to appear most intense. The rules for overtones and binary combinations in the infra-red spectrum are summarized in table 3.

TABLE 3

overtone or combination	symmetry			type of contour
$A_1 + A_2$	s , s , $s \times s$, as , as	s , as , as		inactive
$B_1 + B_2$	as , as , $s \times as$, s , as	s , as , as		inactive
$A_2 + B_2$	s , as , $as \times as$, s , as	as , as , s		C
$A_1 + B_1$	s , s , $s \times as$, as , s	as , as , s		C
$2A_1$	s , s , $s \times s$, s , s	s , s , s		B
$2A_2$	s , as , $as \times s$, as , as	s , s , s		B
$2B_1$	as , as , $s \times as$, as , s	s , s , s		B
$2B_2$	as , s , $as \times as$, s , as	s , s , s		B
$A_2 + B_1$	s , as , $as \times as$, as , s	as , s , as		A
$A_1 + B_2$	s , s , $s \times as$, s , as	as , s , as		A

Inspection of the Raman and infra-red data suggests that some frequencies have rather lower values in the Raman spectrum. In view of the polar nature of the liquid, this is not unusual. Greater significance is therefore to be attached to the infra-red values.

The infra-red band at 605 cm.^{-1} has B -type contour, with the expected spacing of submaxima. It is therefore allocated to class A_1 . This frequency is intense in the Raman spectrum, which agrees with its being connected with a totally symmetrical vibration of this sort. Its value is about that to be expected for a symmetrical vibration of the CCl_2 group, and it is therefore attributed to this mode. In phosgene (Bailey & Hale 1938) the corresponding frequency is 570, in methylene chloride (Corin & Sutherland 1938) about 700, and in thiophosgene (Thompson 1941) 500. The infra-red band at 1620, corresponding to the Raman frequency 1611, has B -type contour, which again implies an A_1 class of vibration. It corresponds to the mode which is largely determined by a vibration of the carbon-carbon double bond. The Raman frequency 295 agrees with what would be expected for a symmetrical deformation of the CCl_2 group. In phosgene its value is 302, in thiophosgene 292, and in methylene chloride 284. The infra-red and Raman frequency of about 3035 is allocated to the symmetrical stretching mode of the CH_2 group. This leaves unallocated one A_1 fundamental, namely, the symmetrical CH_2 deformation.

The infra-red band at 794 has A -type contour, with correct spacing. It is thus a B_2 fundamental, and almost certainly connected with the antisymmetrical vibration of the CCl_2 group. This vibration might be weak in the Raman spectrum. In phosgene there is a similar frequency of 845. The infra-red band at 872 has C -type contour with the expected spacing of submaxima, and there is a corresponding Raman frequency 868. 872 is therefore assigned to a B_1 -type fundamental. The infra-red band at 1094 has a rather distorted though apparently A type contour, and a similar band occurs with vinylidene bromide at about 1060, and with isobutylene at about 1070. Vinylidene chloride shows a weak Raman frequency of 1099. 1094 is therefore assigned to a B_2 fundamental. The antisymmetrical stretching mode of the CH_2 group will have a value of about 3130.

There now remain unexplained Raman frequencies of 375, 450, 540, 1390 and 1558, and a number of weak infra-red bands, and there is one fundamental in each of the four symmetry classes unassigned. The symmetrical CH_2 group vibrations of class A_1 will have a value $1300\text{--}1400\text{ cm.}^{-1}$. The infra-red band at 1318 seems to have A -type contour, and is therefore not of this type. 1391 is therefore assigned to the CH_2 deformation. Now the infra-red band at 970 is explained as a combination ($605 + 375$); 375 is therefore, by the selection rules, either a B_1 or B_2 fundamental. The infra-red band at 1142, with A -type contour, is explained by ($605 + 540$). 540 is therefore a B_2 fundamental, and hence 375 must be the missing B_1 mode. The remaining magnitude 450 is allocated to the twisting motion of class A_2 .

These conclusions are summarized in table 4. This allocation of frequencies gives a satisfactory interpretation of the overtones and combinations observed in the infra-red, as shown in table 5. In this table a few weak bands above 2000 cm.^{-1} have

not been included, since their exact value is uncertain, and in any case several combinations are possible for each of them. The combinations are consistent with the selection rules both in regard to which occur, and also in the band contours. That no other assignment gives such a satisfactory agreement is a further reason for regarding it as probable.

It may be noted that two combinations, namely, $(794 + 295)$ and $(1395 - 295)$, would give absorption close to 1090 cm.^{-1} , and be superposed on the fundamental at 1094. This may partly explain the peculiar contour of this band.

TABLE 4

class	frequencies
A_1	295, 605, 1395, 1620, 3035
A_2	450
B_1	375, 872
B_2	540, 794, 1094, 3130

TABLE 5

infra-red frequency	interpretation	infra-red frequency	interpretation
580	2×295	1580	2×794
714	$1094 - 375$	1742	2×872
829	$375 + 450$	1876	$1094 + 794$
904	(2×450)	1920	$1620 + 295$
914			$1395 + 540$
970	$605 + 375$	1988	$1620 + 375$
1020	$1395 - 375$		$1395 + 605$
1142	$540 + 605$	2070	$(2 \times 605 + 871)$
1318	$450 + 871$		

As regards the unexplained Raman displacements, 1381 might arise from the combination $(872 + 450)$. The frequency 1558 probably occurs as a result of resonance between the first harmonic of 794 and the fundamental of the $\text{C}=\text{C}$ vibration, the latter being rather lower with the liquid than with the vapour.

Two general points should be noted. First, the twisting frequency of 450 falls into line with the values for the same mode in ethylene (950) and vinyl chloride (622), discussed below. Secondly, the absorption bands of vinylidene chloride arising from $\text{C}-\text{H}$ stretching vibrations appear to be unusually weak, as compared with the results found with other molecules containing these linkages.

VINYLDENE BROMIDE

This substance was prepared by adding a cold solution of sodium ethoxide in ethyl alcohol to 1,1,2-tribromoethane. The latter was prepared by adding bromine to vinyl bromide. The vinylidene bromide distilled at 91°C . It polymerized rapidly, so that only the spectrum of the vapour could be measured between 7 and 14μ using

the rapid automatic recorder. The spectrum is shown in figure 3. Table 6 gives the absorption frequencies. Assuming for the approximate molecular dimensions $r_{\text{CH}} = 1.08 \text{ \AA}$, $r_{\text{CC}} = 1.38 \text{ \AA}$, $r_{\text{CBr}} = 1.83 \text{ \AA}$, with angles of 120° , the moments of inertia are about $130, 670$ and $800 \times 10^{-40} \text{ g.cm.}^2$. The molecule therefore approximates fairly closely to being a symmetrical rotator, with $\sigma = 4$ and $S = -0.95$. The

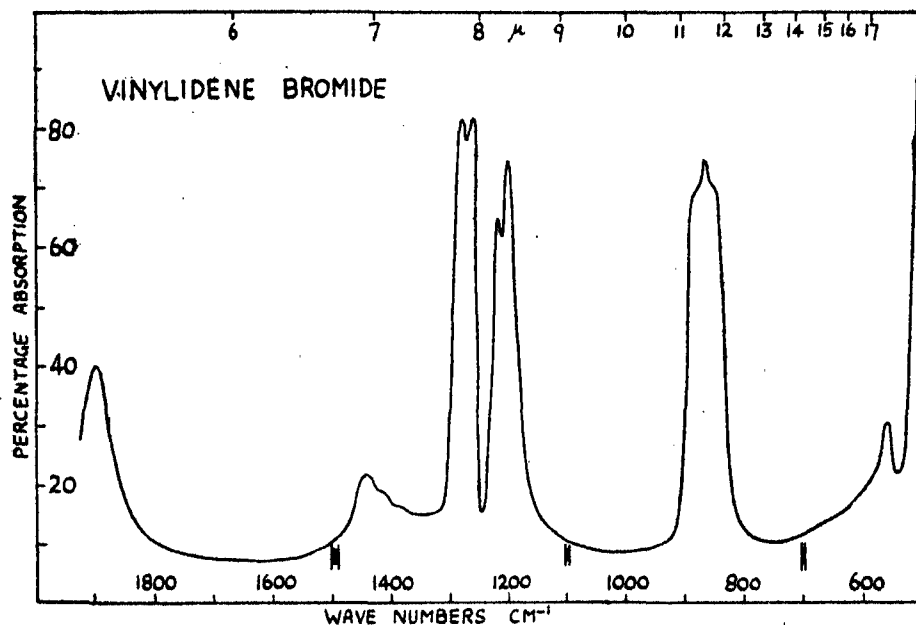


FIGURE 3. Vinylidene bromide. 7 cm. path, 40 mm.

TABLE 6

absorption frequency	intensity
700	strong
730	weak
883	strong
1052 } 1056	strong
1060 }	
1081 } 1085	strong
1088 }	
1171	weak
1400	weak

least axis of inertia will lie, as with vinylidene chloride, in the molecular plane and perpendicular to the $\text{C}=\text{C}$ bond. The *A*-type bands will have a three-branch contour with *P-R* spacing about 9 cm.^{-1} ; the *B*- and *C*-type bands will each consist of a central *Q* branch flanked with shoulders. The twelve normal modes will fall into the same symmetry classes as those of vinylidene chloride. The band at 883 has a perpendicular contour, and is therefore an A_1 or B_1 fundamental; the bands at 1056 and 1085 are *A*-type bands, corresponding to B_2 class vibrations.

The data are meagre, but 883 is probably a B_1 fundamental, a non-planar bending mode, corresponding to 872, with vinylidene chloride. The two magnitudes 1056 and 700 are probably B_2 fundamentals, the former a rocking mode, analogous to 1094 with vinylidene chloride, and the latter the antisymmetrical stretching vibration of the CBr_2 group.

VINYL CHLORIDE

The sample was supplied by I.C.I. (Dyestuffs) Ltd. It was distilled off from the stabilizer before use. The infra-red absorption spectrum of the vapour was measured at various pressures in cells 7 and 21 cm. in length. The spectrum between 3 and 20μ is shown in figure 4, and the positions of the absorption bands are given in table 7. Some of the bands have been located previously by Ta You Wu (1935). Vinyl chloride is a planar molecule, the only element of symmetry being the molecular plane. There will be nine planar modes of vibration, and three non-planar modes, all active in the infra-red and Raman spectra. Roughly, the planar modes can be described as three C—H stretching vibrations, one C=C stretching vibration, one C—Cl stretching mode, and four deformations; the three non-planar vibrations will include one twisting oscillation and two bending motions. Of the four planar deformations, one will approximate to a CH_2 deformation, one to a CH_2 rocking mode, a third to a bending of the C=C—Cl skeleton, and a fourth to a wagging of the odd C—H bond.

According to Brockway *et al.* (1935) it may be assumed for vinyl chloride that $r_{\text{CH}} = 1.08$, $r_{\text{CC}} = 1.38$ and $r_{\text{CCl}} = 1.69$ Å. The angles will each be not very different from 120° . The principal moments of inertia are then about 15, 135 and 150×10^{-40} g.cm.². This means that the molecule is almost equivalent to a symmetrical rotator with moments of inertia 15 and 143×10^{-40} , and least axis in the plane of the molecule and inclined to the C=C bond. Planar vibrations will therefore involve, in general, a change of electric moment with components along more than one axis of inertia, and the contours of these infra-red bands should be hybrid, formed by superposition in definite ratios of types which correspond closely to the parallel and perpendicular types of a symmetrical top. The three non-planar modes will give rise to essentially perpendicular type bands. If the molecule were a true symmetrical rotator with moments $I_A = 15$ and $I_B = I_C = 143$, the spacing between P and R maxima in a parallel band is calculated from the formula of Gerhard & Dennison (1933) to be $17\text{--}18 \text{ cm.}^{-1}$ at room temperatures. In a resolved perpendicular band, the normal spacing of Q branches would be $\frac{h}{4\pi^2} \left(\frac{1}{I_A} - \frac{1}{I_C} \right)$, that is, about 3 cm.^{-1} . Also, in a parallel type band, the central Q branches would be fairly weak.

Many of the observed bands show a doublet character, with spacing $17\text{--}20 \text{ cm.}^{-1}$. These are obviously mainly parallel in type. Other bands, such as those at 895 and 940, have a strong central peak with shoulders on each side, and these are essentially perpendicular type bands, in which the high-frequency shoulders seem to be abnormally intense.

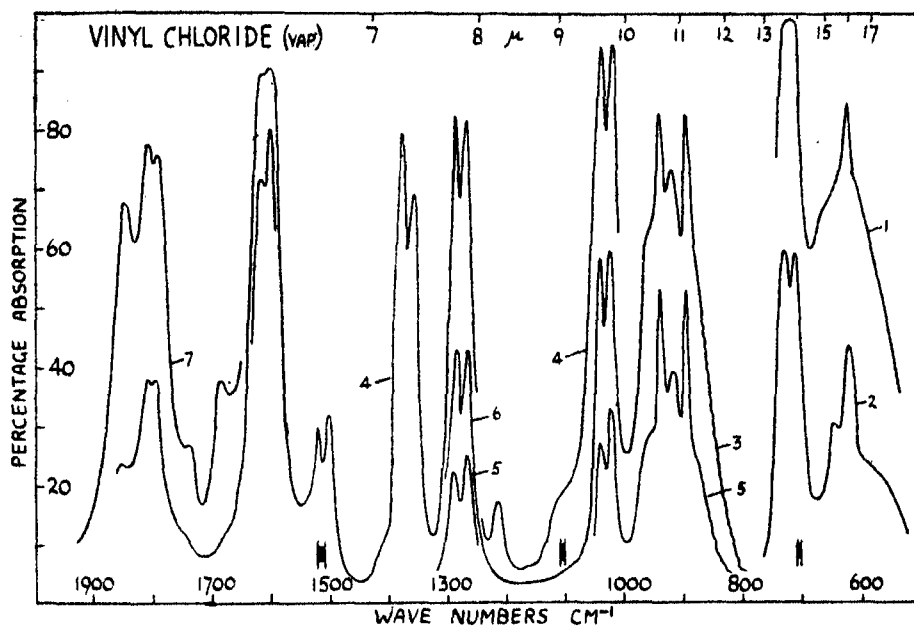


FIGURE 4. Vinyl chloride. (1) 20 cm. path, 300 mm.; (2) 20 cm. path, 50 mm.; (3) 7 cm. path, 200 mm.; (4) 7 cm. path, 450 mm.; (5) 7 cm. path, 100 mm.; (6) 7 cm. path, 200 mm.; (7) 20 cm. path, 780 mm.

TABLE 7

wave number	intensity	contour
622	strong	perpendicular
715 } 724	strong	parallel
732 }		
895	strong	perpendicular
940	strong	perpendicular
1022 } 1030	medium	parallel
1040 }		
1110	very weak	—
1215	very weak	—
1271 } 1280	medium	parallel
1290 }		
1360 } 1370	medium	parallel
1380 }		
1510 } 1520	weak	parallel
1530 }		
1610	strong	—
1686	weak	—
1740	very weak	—
1792 } 1801	medium	parallel
1810 }		
1850	weak	—
2000	very weak	—
2315-2360	medium	—
2645	very weak	—
3096	strong	—

The Raman spectrum has been measured by Pestemer (1930) and by Kohlrausch & Stockmair (1935). With the exception of two doubtful frequencies recorded by Pestemer (1930), the results agree closely, and suggest the following Raman displacements: 396, 620, 703, 908, 1024, 1090, 1242, 1271, 1292, 1360, 1601, 3030, 3078 and 3134. The Raman results fall into line with the infra-red measurements, and suggest values for all twelve normal modes, as given in table 8. By analogy with related molecules, 724 is assigned to the vibration which mainly involves stretching of the C—Cl bond, and 1620 to a similar vibration of the C=C linkage. The three essentially parallel type bands at 1370, 1030 and 1280 are attributed to the CH_2 planar deformation, a CH_2 planar rocking mode and the C—H wagging. Similar frequencies at 1360, 1010, 1260 are found with vinyl bromide as explained below. The low frequency 395 is attributed to the planar deformational motion of the C=C—Cl skeleton. If the three magnitudes 3030, 3130, 3080 are assigned to stretching motions of the C—H bonds, all the nine planar modes have been determined. Of the three values 622, 897 and 940, 897 can be assigned to a bending mode similar to that found with isobutylene at 890, and 940 to the other bending mode. 622 is then the twisting frequency.

TABLE 8

fundamental frequency	type	infra-red combinations		Raman combinations	
395	δ_{CCl}	1110	395 + 724	1242	2×620
622	twisting	1215	1610 - 395	1292	395 + 908
724	ν_{CCl}		—	—	—
895	δ	1430?	(2×724)	—	—
940	δ		(1030 + 395)	—	—
1030	δ		—	—	—
1280	δ	1520	622 + 895	—	—
1370	δ_{CH_2}		—	—	—
1610	ν_{CC}	1686	(1280 + 395)	—	—
3030	$\nu_{\text{CH}_2}^s$		(1030 + 622)	—	—
3130	$\nu_{\text{CH}_2}^a$		(724 + 940)	—	—
3080	ν_{CH}	1740	1030 + 724	—	—
			1370 + 395	—	—
		1801	2×895	—	—
		1850	(895 + 940)	—	—
			(2×940)	—	—
		2000	1280 + 724	—	—
		2315-2360	(1610 + 724)	—	—
			(1370 + 940)	—	—
		2645	(1280 + 1370)	—	—
			(1610 + 1032)	—	—

The combinations in the infra-red and Raman spectra are also satisfactorily interpreted in table 8.

The shoulders of the band at 895 are just resolved into a succession of Q branches, the spacing of which is close to 4 cm.^{-1} . These are shown in figure 5, and table 9 gives the positions of the Q branches measured around the central peak at 895 cm.^{-1} ,

although they could not be measured with high precision. The spacing of 4 cm^{-1} found is rather greater than that expected, but the difference may be explained either by interactions of the kind common to perpendicular type oscillations, or by small inaccuracies in the model dimensions assumed above.

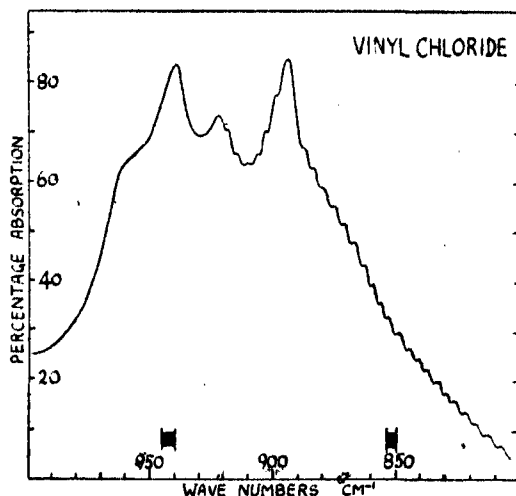


FIGURE 5. Vinyl chloride. Perpendicular bands.

TABLE 9

wave number of Q branch	spacing	wave number of Q branch	spacing	wave number of Q branch	spacing
923	—	880	4	841	4
919	4	876	4	837	4
915	4	872	4	833	4
911	4	868	4	829	4
908	3	864	4	826	3
904	4	861	3	822	4
900	4	857	4	818	4
895	centre	853	4	813	5
888	—	849	4	808	5
884	4	845	4		

VINYL BROMIDE

This substance was prepared by treating ethylene dibromide with alcoholic caustic potash as described by Swarts (1901). The product distilled at 16°C . Its very rapid polymerization limited the measurements which could be made. Figure 6 shows the absorption spectrum between 5 and 20μ . Table 10 gives the position of the absorption maxima in cm^{-1} . The molecule can be assumed to be planar, with $r_{\text{CH}} = 1.08\text{ \AA}$, $r_{\text{CBr}} = 1.83\text{ \AA}$, $r_{\text{CC}} = 1.38\text{ \AA}$, and angles of about 120° . The moments of inertia are then about 17 , 192 and $209 \times 10^{-40}\text{ g.cm}^2$. Thus, like vinyl chloride, the

molecule approximates closely to being a symmetrical rotator, with the least axis oblique to the C=C bond. The spacing between *P* and *R* maxima in a parallel-type band will then be about 16 cm.^{-1} . The perpendicular bands will have a strong central maximum, and, if the *Q* branches are resolved, their separation will be about 3 cm.^{-1} , as with vinyl chloride.

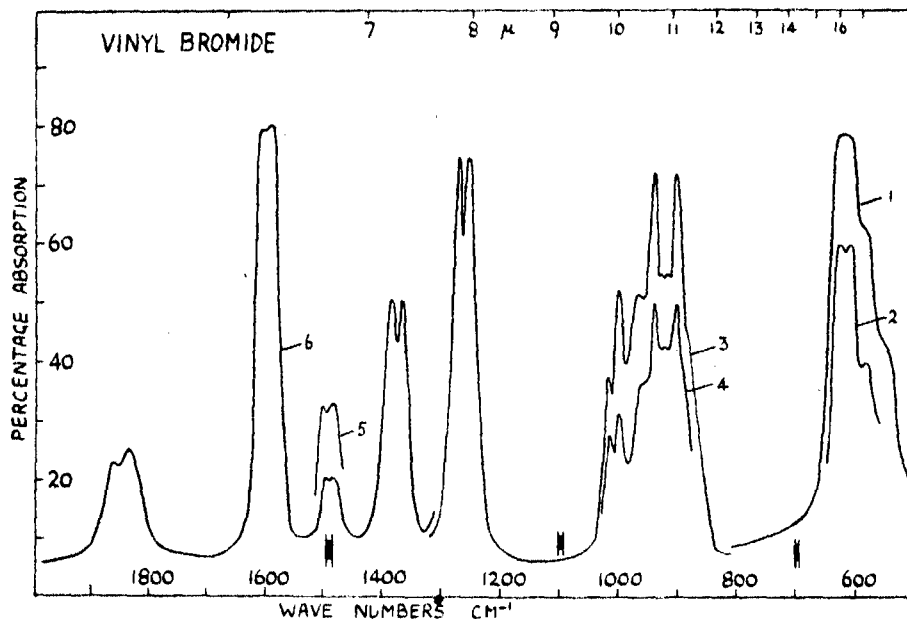


FIGURE 6. Vinyl bromide. (1) 20 cm. path, 100 mm.; (2) 20 cm. path, 25 mm.; (3) 7 cm. path, 50 mm.; (4) 7 cm. path, 25 mm.; (5) 7 cm. path, 350 mm.; (6) 7 cm. path, 200 mm.

TABLE 10

position of band (cm.^{-1})	type	interpretation
552	---	902-345
587	---	940-345
608 } 622 }	—	fundamental
902	perpendicular	fundamental
940	perpendicular	fundamental
1002 } 1016 }	parallel	fundamental
1255 } 1270 }	parallel	fundamental
1370 } 1385 }	parallel	fundamental
1484 } 1500 }	? parallel	902 + 605
1595 } 1615 }	? parallel	fundamental
1837	---	940 + 902
1865	---	1260 + 605

The Raman spectrum has been measured by Bourguet & Piaux (1935) and by Kohlrausch & Stockmair (1935). The most reliable values for the displacements are 345, 497, 600, 900, 1000, 1245, 1369, 1598, 3014, 3076 and 3103. A comparison of these values with the infra-red data, and a consideration of the rotational contours of the infra-red bands, suggest for the fundamental frequencies the values given in table 11. The Raman frequencies which were determined, using the liquid, seem as a rule to be rather smaller than the corresponding infra-red frequencies. The combinations observed in the infra-red spectrum are also interpreted in table 10. The Raman displacement of 1245 can be explained by $(902 + 345)$, although it might be connected with the 1262 fundamental. The abnormal contour of the absorption at 1002 might be due to overlapping of the fundamental at 1008 with a harmonic of 497.

TABLE 11

fundamental frequency	type	fundamental frequency	type	fundamental frequency	type
345	δ_{CBr}	940	δ	1605	ν_{CO}
497	twisting	1008	δ	3014	ν_{CH_2}
615	ν_{OBr}	1262	δ	3100	$\nu_{\text{OH}_2}^a$
902	δ	1377	δ_{CH_2}	3076	ν_{CH}

As with the corresponding band of vinyl chloride, the band of vinyl bromide at 902 just shows at the sides a resolved succession of Q branches with spacings of 3–4 cm^{-1} . The approximate positions of the Q branches on the lower frequency side of the band centre are given in table 12.

TABLE 12

wave number of Q branch	spacing	wave number of Q branch	spacing	wave number of Q branch	spacing
894	—	872	4	850	3
890	4	869	3	846	4
887	3	865	4	840	6
883	4	861	4	835	5
880	3	857	4	831	4
876	4	853	4		

CORRELATION OF VIBRATION FREQUENCIES OF VINYL HALIDES

Although the infra-red spectrum of vinyl iodide has not been measured (see note below), the Raman spectrum has been recorded by Kahovec & Kohlrausch (1940), who found the following displacements: 309, 435, 535, 879?, 982, 1098, 1229, 1369, 1581, 3003, 3062 and 3092. If these values are compared with those given above for the other vinyl halides, correlations can be made (figure 7) which suggest that, with the exception of the two weak displacements 879 and 1098, all the remaining Raman frequencies are due to fundamentals. Ten normal vibration frequencies are thereby fixed. The remaining pair will lie between 900 and 1000, but their exact values

cannot be determined. It should also be remembered that the Raman frequencies adopted for liquid vinyl iodide may be rather smaller than would be found with the vapour in the infra-red spectrum.

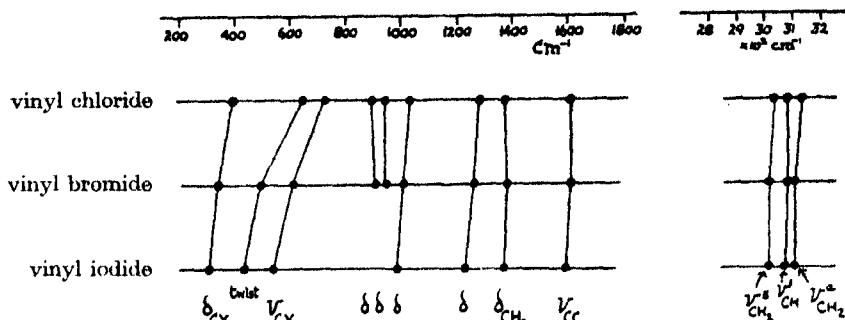


FIGURE 7. Correlations with the vinyl halides.

CHLOROPRENE

This substance was originally examined, since it is the parent monomer of the neoprene group of polymers. It has other obvious relationships with isoprene and the vinyl compounds discussed above. The substance was supplied by I.C.I. (Dye-stuffs) Ltd. as a solution in xylene, stabilized with a little catechol. It was distilled from the solution just before use. Most of the measurements were made with the vapour, in a cell 7 cm. in length, with various pressures. The absorption curves are shown in figure 8. The spectrum of a thin film of liquid was also measured. Polymerization occurred so slowly as to cause no difficulties in these measurements.

The positions of the absorption maxima with the vapour are given in table 13, and, as can be seen from the curves of figure 8, some of the bands show a well-marked contour. The Raman spectrum has been measured by Kubota (1938) who found the following displacements: 158, 249, 387, 518, 629, 735, 882, 923, 1020, 1214, 1287, 1359, 1381, 1417, 1526, 1581, 1628, 3015 and 3113. The agreement between many of these values and the positions of the infra-red bands is close. Raman measurements have also been recorded by Perymova (1940), using mixtures of chloroprene and its polymer. These results confirm many of the above values.

Chloroprene has twenty-four normal modes, and all will be permitted to appear in both the Raman and infra-red spectra. Exact geometrical description of all the vibrations is impossible, but some approximate correlations may be useful. Thus, as a first approximation, the twenty-four normal modes may be divided into two

groups, nine involving motions of the skeleton $\begin{array}{c} \text{C}-\text{C}-\text{C}-\text{C} \\ | \\ \text{Cl} \end{array}$, and fifteen involving

movements which are localized in the C—H bonds. The nine modes of the skeleton will include one vibration largely determined by a vibration of the C—Cl bond, one by the C—C bond, two by C=C bonds, and there will be three planar deformations

and two non-planar deformations. The other fifteen modes can be described as five C—H stretching vibrations, and ten oscillations which involve deformation, rocking or bending of the CH_2 and CH groups. These ten oscillations will be roughly described as two CH_2 planar deformations, two planar CH_2 rocking motions, one planar CH rocking, two non-planar CH_2 bending motions, one non-planar CH bending, and two twisting modes.

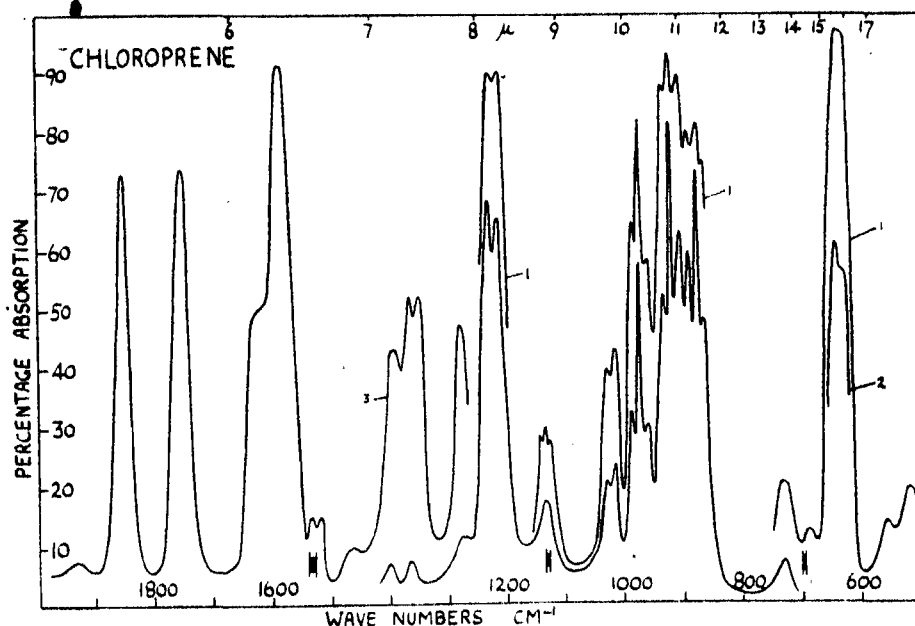


FIGURE 8. Chloroprene. (1) 7 cm. path, 120 mm.; (2) 7 cm. path, 50 mm.; (3) 20 cm. path, 135 mm.

By analogy with its value in related molecules, the C—Cl stretching mode will be expected to have a value 600–750 cm^{-1} . The band at 738 is weak, and it is more satisfactory to assign that at 640 to this mode. 1590 and 1620 are probably connected with the vibrations of C=C bonds, the lower value being associated with that motion which is more determined by the C=C bond to which the chlorine is attached. The five C—H stretching modes will have values 3000–3200.

Bending motions of the skeleton are less easily specified, but in the previous paper it was shown that in hydrocarbons the vinyl radical $\text{CH}_2=\text{CH}-$ gives rise to two frequencies close to 909 and 990. In vinyl chloride these two vibrations lie at 895 and 940. Similarly, the radical $\text{CH}_2=\text{C}<$ in hydrocarbons such as isobutylene gives rise to a single frequency close to 890; in vinylidene chloride it was shown above to be 872. In chloroprene the three corresponding modes will have the values 924, 974 and 878. There is no Raman displacement corresponding to the infra-red band at 974. It may be noted that isoprene has three similar bands at 907, 994 and 895. The

peculiar contour of the infra-red band of chloroprene at 1025 suggests that it involves a rocking motion of the CH_2 group, corresponding to 1090 with vinyl chloride and 1060 with vinyl bromide, and 1075 with isoprene.

TABLE 13

position of band (cm^{-1})	intensity	position of band (cm^{-1})	intensity
524	very weak	1131	
559	very weak	1137	weak
635		1143	
645	strong	1218	
697	very weak	1232	1225 strong
738	very weak		1285 medium
867		1354	
878	strong	1370	1362 medium
890			1400 medium
907			1471 very weak
924	strong	1524	
934		1538	1531 medium
962			1590 strong
974	medium		1620 medium
986			1757 strong
1020			1852 strong
1025	weak		1930 weak
1034			3030 medium
			3120 medium

The four remaining π -skeleton deformations may have values 158, 249, 387, the lowest Raman displacements, and 559, a weak infra-red band; the second planar CH_2 deformation may be 1400, an infra-red frequency, and the second planar CH_2 rocking mode another infra-red frequency of 1137. The remaining twisting frequencies may be 697, a weak infra-red band.

The array of fundamentals would then be: 158, 249, 387, 524, 559, 640, 697, 738, 878, 924, 974, 1025, 1137, 1225, 1285, 1362, 1400, 1590, 1630 and five values close to 3000. While some of these assignments are tentative, all seem reasonable, and give a satisfactory interpretation of other Raman and infra-red combination frequencies. Thus the Raman frequency 1526 could be interpreted as either $(974 + 559)$ or $(640 + 878)$, and the occurrence of two close frequencies, 1381 and 1417, might arise from superposition of a fundamental 1400 with the harmonic of 697, leading to a resonance splitting. Infra-red bands at 1471, 1531, 1757, 1852 and 1930 can be explained by (2×738) , $(974 + 560)$, (2×878) , (2×924) and (2×974) .

It is interesting to consider how far the infra-red data help to decide between the *cis* and *trans* planar forms for chloroprene. Assuming approximate values for the molecular dimensions, namely, $r_{\text{CCl}} = 1.69 \text{ \AA}$, $r_{\text{C-C}} = 1.46 \text{ \AA}$, $r_{\text{C=C}} = 1.35 \text{ \AA}$, $r_{\text{CH}} = 1.08 \text{ \AA}$, and angles of 120° , the *trans* form has approximate moments of inertia $150, 220$ and $370 \times 10^{-40} \text{ g.cm}^2$, and the *cis* form $90, 320$ and 410×10^{-40} .

The band contours will generally be hybrid, obtained by compounding those of the *A*, *B* and *C* types, in which the change of electric moment is parallel to the least,

middle and greatest axes of inertia. Using Badger & Zumwalt's notation, the *trans* form will have $\sigma = 0.85$ and $S = 0$. The *A*-type bands will show three submaxima, the spacing between the outermost of which will be $14\text{--}15\text{ cm.}^{-1}$; the *B*-type bands will have essentially a double maximum type of contour, with spacing about 9 cm.^{-1} ; and the *C*-type bands, in which the change of electric moment is perpendicular to the molecular plane, will have a strong central peak with weaker submaxima on each side of it, the spacing of the two outer maxima being about 25 cm.^{-1} . The *cis* form approaches more closely to being a symmetrical rotator, with $\sigma = 2.8$ and $S = -0.85$. The *A*-type bands will have a spacing of outermost maxima of about 14 cm.^{-1} ; the *B*- and *C*-type bands will not be expected to show a very marked central peak, and the spacing of the outermost maxima should not exceed about 15 cm.^{-1} .

The experimental results on the band contours are seen to be in much better agreement with the *trans* form than with the *cis*. The *C* (perpendicular) type bands at 878, 924 and 974 have just the contour to be expected, and although obviously complicated by being hybrids, other spacings of 12, 14, 16 and 14 are in rough agreement with essentially *A*-type bands. Moreover, with the *trans* structure the least axis of inertia is inclined to the C—Cl bond at a markedly greater angle than in the *cis* form. This means that if the *trans* structure predominates, the vibration band largely determined by the stretching of the C—Cl bond will have a more nearly *B*-type contour. The observed spacing of the fundamental at 640 agrees with this.

The infra-red spectrum of thin layers of liquid chloroprene agrees in all essentials with that of the vapour, but the relative intensities and exact frequencies of some of the bands are altered, and a few other feeble bands are observed. These differences may be caused by modification of the absorption intensities as a result of intermolecular influences in the liquid, and they will be more conveniently discussed elsewhere.

THE SPECTRA OF POLYVINYL CHLORIDE, POLYVINYLIDENE CHLORIDE, HALOTHENES AND NEOPRENES

Polyvinyl chloride was supplied by I.C.I. (Dyestuffs) Ltd., and polyvinylidene chloride by I.C.I. (Plastics) Ltd. Films of the former were made by evaporation of solutions in ethylene dichloride from the surface of mercury, and of the latter by rapid evaporation of solutions in pentachloroethane from a heated glass plate, and of a dilute solution in boiling dioxane. Care was taken to remove all the solvent, and this was verified as far as possible by the absence of its main absorption bands. It was sometimes found impossible to remove all the solvent from a film, even after heating *in vacuo*. Excessive heating also sometimes appeared to lead to changes in spectrum which implied that structural changes had taken place.

The spectra of the polymers between 6 and 20μ are shown in figure 9, the films being about 0.05 mm. thick. They show well-marked differences from each other, and also from those of the parent monomers given above. Several interpolymers of

vinyl chloride with vinylidene chloride have also been measured. The spectra of such interpolymers are not merely a superposition of the main features of the separate polymers, but contain new bands not present in either of the latter, as might be expected since fresh units in the nuclear skeleton may arise.

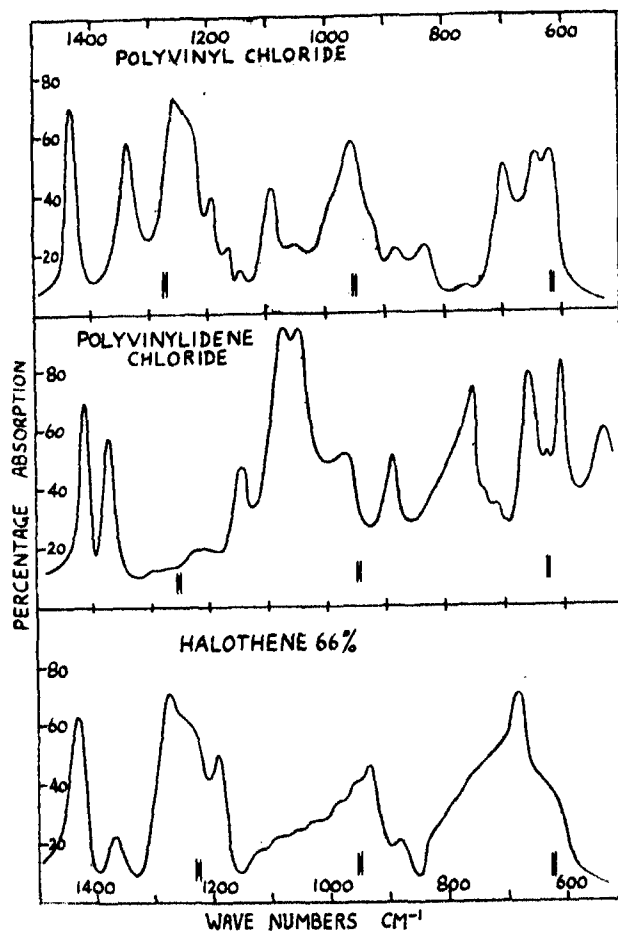


FIGURE 9. Polyvinyl chloride, polyvinylidene chloride and halothene.

The spectra of a series of polychloroprenes (neoprenes) have also been measured and show interesting differences. Correlation of the observed absorption frequencies with particular motions of the nuclear skeleton can only be attempted by comparison of the spectra with those of other compounds of related structure. With this object, the spectra have been measured of (a) a series of halothenes (chlorinated polythenes) in which the percentage of chlorine was gradually increased from zero to about 70 % by weight, and (b) a number of simple chlorinated hydrocarbons of low molecular weight and containing different types of nuclear skeleton. The spectrum of a sample of halothene containing 66 % chlorine is shown; it shows some resemblances

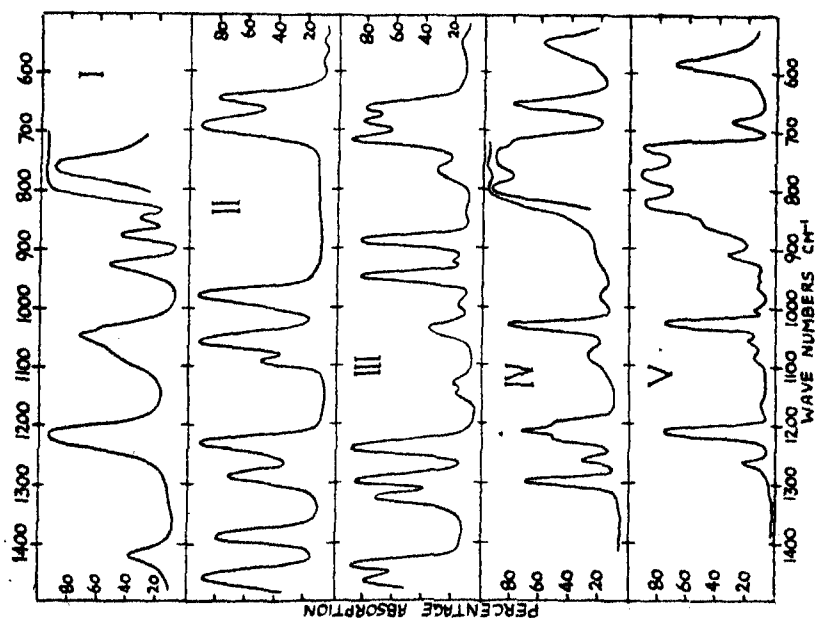


FIGURE 10a. I, chloroform; II, ethylidene chloride; III, ethylene dichloride; IV, acetylene tetrachloride; V, pentachloroethane.

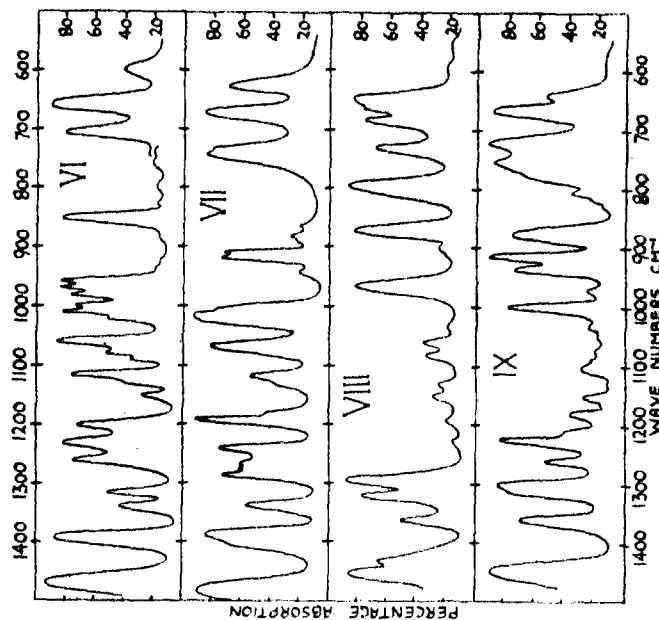


FIGURE 10b. VI, 2,3-dichlorobutane; VII, 2,3-dichloropropane; VIII, 1,3-dichloropropane; IX, 1,2,3-trichloropropane.

with those of the other polymers, and some marked differences, but the data are not yet complete enough to discuss in detail.

Figure 10 shows the spectra of some of the simple chlorinated paraffins so far examined. These were:

Chloroform: a repurified commercial product, b.p. 62°C .

Ethylene dichloride: redistilled B.D.H. product, b.p. 83.5°C .

Ethylidene chloride: redistilled B.D.H. product, b.p. 59.2°C .

Acetylene tetrachloride: sample from the Dyson Perrins laboratory, redistilled, b.p. 147°C .

1.3-Dichloropropane: prepared by heating trimethylene glycol with fuming hydrochloric acid at 100°C in a pressure bottle. The sample boiled at 125°C .

1.2-Dichloropropane: prepared by the direct addition of chlorine to propylene in the gas phase. The sample boiled at 98°C .

Pentachloroethane: sample from the Dyson Perrins laboratory, redistilled, b.p. 159°C .

1.2.3-Trichloropropane: prepared by adding chlorine to allyl chloride. The sample boiled at 158°C .

2.3-Dichlorobutane: sample from the Dyson Perrins laboratory, b.p. 118°C .

The spectra of each of these substances were measured in a cell about 0.05 mm. in thickness. The curves reveal marked differences, and at first sight there are no obvious correlations between them. A closer examination suggests certain regularities, however, and when more related substances have been examined these will be considered in detail.

Note added in proof. Since submitting this paper, we have been able to measure the spectra of vinyl iodide and vinyl fluoride, and to complete the correlation of the vibration frequencies of all the vinyl halides and vinyl cyanide. These results have been published in *J. Chem. Soc.* (1944), pp. 303, 597 and *Trans. Faraday Soc.* (1945), May.

We are greatly indebted to I.C.I. (Dyestuffs) Ltd., I.C.I. (Alkali) Ltd., and I.C.I. (Plastics) Ltd., for samples of the substances quoted above, and for their interest in this work. We also thank the Royal Society and the Chemical Society for financial grants.

REFERENCES

- Badger & Zumwalt 1938 *J. Chem. Phys.* **6**, 711.
Bailey & Hale 1938 *Phil. Mag.* **25**, 98.
Bourguet & Piaux 1935 *Bull. Soc. Chim. Fr.* (5), **2**, 1958.
Brockway, Beach & Pauling 1935 *J. Amer. Chem. Soc.* **57**, 2693.
Cabannes 1938 *J. Chim. phys.* **35**, 1.
Corin & Sutherland 1938 *Proc. Roy. Soc. A*, **165**, 43.
Emschwiller & Lecomte 1937 *J. Phys. Radium*, **8**, 130.
Gerhard & Dennison 1933 *Phys. Rev.* **43**, 197.

- Kahovec & Kohlrausch 1940 *Z. phys. Chem. B*, **46**, 165.
Kohlrausch & Stockmair 1935 *Z. phys. Chem. B*, **29**, 292.
Kubota 1938 *Bull. Chem. Soc. Japan*, **13**, 678.
Perymova 1940 *J. phys. Chem. U.S.S.R.* **3**, 346.
Pestemer 1930 *Mh. Chem.* **57**, 469.
Swarts 1901 *Chem. Zbl.* [5], **5**, 804.
Ta You Wu 1935 *Phys. Rev.* **46**, 465.
Thompson 1941 *Trans. Faraday Soc.* **37**, 251.
-

The topography of crystal faces

I. The topography of a (100) face of a left-handed quartz crystal

By S. TOLANSKY, *University of Manchester*

Communicated by Sir Lawrence Bragg, F.R.S. —Received 14 April 1944

[Plate 1]

A description is given of a multiple beam interferometric procedure which can be used as a powerful method for revealing the details of the surface topography of approximately flat crystal planes. The optical conditions are critical and are discussed. Details of structure are revealed which cannot be detected by the microscope or the goniometer. The interference fringes which are employed can be used either in transmission or in reflexion, the latter type being well suited for the examination of translucent or opaque crystals. Face angles, between small faces or large faces, which are no more than 1/50th of a minute of arc can be measured.

The topographical features of a highly lustrous (100) face of a left-handed quartz crystal are studied by this procedure. Large vicinal faces, inclined at angles varying from 0.50 to 9.00 min. of arc, are found to cover the (100) face. The surfaces of most of these vicinal faces are curved, with radii of curvature varying from 20 to 60 m. The angles of some of the vicinal faces in contact are found to vary considerably along the length of the common edge.

The characteristic visual markings of a (100) face, namely striations, \wedge marks and triangles, reveal their influence upon the fringe patterns and their topographical features are determined. The striations are shown to be either minute ridges or ruts, the height or depth being only 100 Å (20 silica molecules). The \wedge markings reveal small discontinuities in level in their neighbourhood. The triangular markings are shown to be submicroscopic tetrahedral projections of some 450 Å in height (90 molecules), with vicinal face angles of the order of 2 or 3 min. The nature of the markings is discussed.

The experimental method described can be considered to function in a region intermediate between the microscope and X-rays.

INTRODUCTION

In standard optical practice, the optical figure of a plane glass surface is frequently examined by matching against a flat, using Fizeau's fringes, in reflexion. It was pointed out as long ago as in 1897 by Fabry and Perot that a very great increase

in both precision and in brilliance arises if multiple beams are employed. This is achieved by depositing high-reflecting-coefficient transmitting silverings on the surfaces. With correct optical conditions the resulting multiple reflexions lead to the production of extremely sharp localized fringes which have an intensity distribution closely resembling the characteristic sharpened distribution of the rings of equal inclination at infinity given by the Fabry-Perot interferometer. These fringes are most frequently viewed in transmission, in which case brilliant fine lines, on a broad dark background, can be seen with a suitable source.

This modification of Fizeau's fringes has been used for examining the ultimate perfection of the best worked optical flats, and with it, errors in fringes of less than a hundredth of a wave can readily be detected. In this paper it will be shown how these fringes, both in transmission and reflexion, can be used as a powerful, easily applied, optical weapon for revealing sub-microscopic detail in the surface topographical features of approximately plane crystal surfaces. The method described is capable of wide application and is a new method of approach. The precision which can be attained is very high indeed, for under the most favourable conditions displacements of only $\lambda/200$ can be measured, and with the green mercury line as a source this is a matter of some 30 Å only. Thus molecular dimensions can be approached and the procedure can be regarded as bridging a gap between the X-ray and the microscopic study of crystals. If intensively applied it should provide information about crystal growth, crystal strength, cleavage behaviour and other aspects in which the nature of the crystal surface is important.

In this paper the experimental procedure is described and the method applied to the examination of the topographical features of a (100) face of a quartz crystal. The optical conditions for the production of very sharp fringes are critical and will be considered in detail first.

MULTIPLE BEAM INTERFEROMETRY WITH CRYSTAL SURFACES

The experimental procedure is very simple. It is necessary to work with a crystal surface which is approximately plane. A high-reflecting-coefficient thin film of silver is deposited upon the crystal surface by evaporation *in vacuo*, the reflecting coefficient for the green mercury line being of the order of 0.87. (The technique for the deposition of high grade silverings had formerly been developed for the preparation of Fabry-Perot interferometer mirrors.) It is necessary to have (a) a good vacuum, (b) a very clean crystal surface, (c) a highly uniform deposit. With crystalline quartz surfaces, the only mechanical cleaning used was light rubbing with cotton wool soaked in hydrogen peroxide solution. The main cleansing mechanism used was ionization de-gassing by a glow discharge, immediately prior to evacuating hard for the purpose of deposition by evaporation of the silver film.

The silver film evaporated on to the crystal surface is some 500 Å thick. The necessary film thickness can be estimated, after some practice, by the transmission

colour. The silver cloud deposits itself so as to contour *exactly* all the detail on the crystal surface. That this is true to within the limits of measurement, namely 30 Å, is proved by the observation on mica, recorded in part II. It is necessary to add that this argument need not be considered to apply to re-entrant surfaces, which are in any event not accessible for observation by the procedure described here. The silvered crystal is mounted on a screw-controlled bed and brought into close proximity with a high grade optical flat similarly coated with silver. The particular flat employed was a 5 cm. quartz disk accurately worked to be correct *over the whole surface* to $\lambda/20$. The existing error was largely a matter of general curvature, and by matching against a flat accurate to $\lambda/60$ it was clear that over moderately small areas the sum of the errors on both flats was less than $\lambda/100$. The inevitable errors on the flat have no influence upon the conclusions reached, for the same topographical features appear when the crystal is matched against different portions of the flat. This clearly proves that they are to be attributed to the crystal only.

The crystal surface and the optical flat are brought into almost exact parallelism and as close together as is possible. By this means thin film localized fringes either in transmission or in reflexion can be produced. The formation of such fringes is not dependent upon there being a thin film between the silvered surfaces but the *sharpness* is considerably affected by the separation. It is in fact possible to produce the Fizeau type of fringes with silvered flats as far as 5 cm. apart, but the fringe visibility is then very poor indeed. The necessary conditions for obtaining very sharp fringes will now be considered.

The interference fringes formed in a thin film are of course governed by the well-known expression $n\lambda = 2\mu t \cos \theta$ in which the symbols have the usual meaning. If the angle of incidence, θ , is kept constant, then the fringes which form are true contour lines, representing true fringes of equal thickness. In the case of ordinary (unsilvered) thin film fringes, in which *only two* light beams are involved, moderate variations in θ are of no consequence, because the intensity distribution of the fringes takes on a \sin^2 form. Attention has apparently not been directed to the fact that this condition no longer applies when multiple beams are employed. The summing up of the multiple beams leads to a great fringe sharpening, and in fact so narrow can they be, that to obtain the best results one cannot tolerate much more than a 0.1 % change in the value of $\cos \theta$. Such a change, small as it is, makes a 1 % change in the order of interference of the 10th order fringe and such a change is a drawback in so far as a 1 % change in order leads to a broadening which can just be detected. If then the incidence is close to normal the tolerance permitted in θ does not exceed more than 2–3°.

It is thus essential to arrange to collimate carefully the incident light so as to obtain a reasonably parallel beam. This is most simply achieved by employing a small source at the focus of a lens. The diameter of the source was approximately 1 mm. when employing a spherically corrected lens of 15 cm. focal length, a smaller source leading to an unnecessary loss in light.

A second, and much more important feature differentiating multiple beam from simple two beam interference, is that in the former case the incidence must be restricted to normal and at the same time the separation between the interference faces must be as small as possible, if high grade definition is to result. Although it is true that normal incidence is often employed in the classical method of viewing Fizeau fringes, and in the more recent method of viewing the transmitted silvered modified fringes, both the explanation of the necessity for normal incidence, and the restriction as to separation, seem to have escaped attention. The explanation is as follows. The fringes, when incidence is normal, are localized in the interference film and are viewed there. For the fringe sharpening due to the multiple beams to be fully effective it is necessary to integrate *locally* all the interfering multiple beams. With fringes of equal inclination, at infinity (Fabry-Perot fringes) this is automatically achieved by the collecting lens. With the localized wedge fringes now under consideration, the necessary condition is that the *lateral* displacement, along the surface, of each successive multiply reflected beam contributing to a given fringe, should only be a very small fraction of the interval between orders. With the high reflecting coefficients employed here, the 50th image has still appreciable intensity, and if this is not to be displaced by more than, say, 1/100th of an order, then a simple calculation shows that at normal incidence, with say two fringes to the millimetre (a convenient dispersion) the separation of the surfaces must be less than 1/20th of a millimetre. For incidence other than normal the separation must be progressively reduced.

Thus it is clear that to obtain high grade definition the collimation must be exact. Not only must the incident light consist of a parallel beam, it must also fall normally on to the interference film, which must be as thin as possible.

A suitable source is a vacuum mercury arc and for most general purposes this can be filtered crudely to pass either the green, or the green and yellow lines. Filters can frequently be dispensed with, for particular exposures. Both transmitted and reflected fringe systems can be employed, according to the particular problem and in what follows their application will be separately considered.

THE TOPOGRAPHY OF A (100) FACE OF A QUARTZ CRYSTAL

The specimen of quartz selected for study, a left-handed crystal, is shown diagrammatically in figure 1. This crystal possesses one large rhombohedral face; in the notation of Miers it is $R(100)$. The crystal was chosen because this particular face appeared to show a very high degree of optical perfection when viewed by simple reflexion. The lustre was brilliant, reflected images did not show distortion, and no multiple Brewster images could be detected. Some characteristic (100) face markings could be seen with a lens, providing light of critical incidence was used to illuminate the surface. The details of the visible markings are shown in figure 1. The (100) face is strongly outlined and the barely visible surface markings are rendered in lines which are heavier for the more easily noticed features. The

correlation of these surface markings with the interference patterns is discussed in detail below.

The (100) face was silvered and mounted against the optical flat. The transmitted fringe system will first be considered.

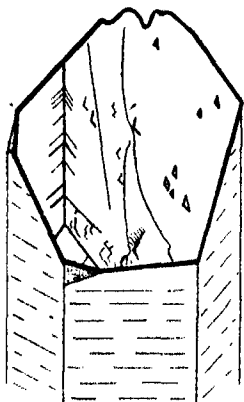


FIGURE 1. Quartz (100) face, visual markings.

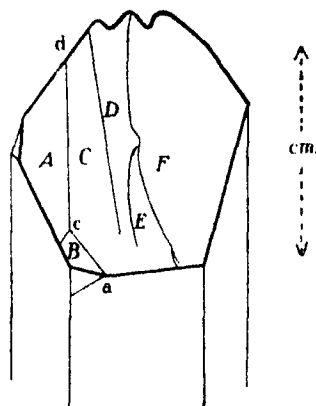


FIGURE 2. Quartz (100) face.

(a) Transmitted fringes

Accurate collimation presented difficulties. The light enters from behind, at one of the long rectangular prism faces and must be incident at an acute angle such that it refracts to strike the inclined rhombohedral face at normal incidence. The face on which the light is incident showed the characteristic coarse striations exhibited by such quartz faces. This rough entry effectively destroyed critical collimation. This serious defect was overcome by attaching to the face of entry a thin microscope cover slide, Canada balsam being used as the attaching medium. The refractive index of the balsam is sufficiently close to that of quartz to remove effectively all the striations. The improvement in definition consequent on critical collimation was striking.

As the beam which manages to reach the rhombohedral face (100) is necessarily limited in aperture by the rectangular shape of the first refracting prism face, the whole of the (100) face is not illuminated. In order to cover the face completely it was necessary to resort to the reflected system, in a manner to be described later.

Two typical transmission contours are shown in figures 3 and 4 (plate 1). The interpretation is materially assisted by the employment of such sets of contours taken with different angles of inclination between the crystal face and the flat. The two figures were obtained by tilting the crystal face differently, and though superficially they appear to differ it will be recognized that they are in fact different viewpoints of the same surface configuration. Figure 3 corresponds to the type of contour map familiar in geography. This is obtained by intersecting the surface features with equidistant horizontal planes. Figure 4 shows the contour lines resulting from the same surface when the intersecting equidistant planes are inclined to the horizontal.

Attention may be drawn to a number of special features.

(1) The extreme sharpness of the fringes is clearly evident. It is apparent that displacements which are very small fractions of an order, *over quite small areas*, can readily be detected and accurately measured. The area of the crystal surface is about 1 sq.cm.

(2) Apart from the general major contour line distribution, a great deal of minor fine detail is rendered. This is almost entirely attributable to the crystal, and not due to imperfections of the optical flat.

(3) By altering the tilt of the crystal face and observing the direction of movement of the fringes, it can be decided without ambiguity which areas are elevated and which are depressions.

(4) It is to be noted that the magnification is only enormous in the direction perpendicular to the surface of the flat.

Before reviewing the interpretation of the contours, the reflected fringe system will be discussed.

(b) Reflected fringes

Hamy first showed, in 1906, that narrow localized fringes, resembling reversed transmission fringes are given by reflexion from a silvered wedge, providing the reflecting coefficient is sufficiently high. With the reflecting coefficient employed here the back reflected fringes are complementary to the transmitted system (this is not always the case with low reflecting coefficients). The fringes thus appear as fine *dark* lines upon a broad bright background. An important distinction between the transmitted and reflected systems is that for the latter the source must closely approximate to monochromatism. Whereas in transmission two distinct wave-lengths appear as separate (coloured) fringe patterns, which may or may not occasionally show overlapping of a fringe here and there, in reflexion, the intense light 'background' of the one system superposes on the dark 'absorption fringes' of the other and reduces the visibility to such an extent that even with the few wave-lengths emitted by an unfiltered mercury arc, the fringes almost completely vanish. This defect is of little consequence if a suitable filter is employed.

If the reflected system is viewed through a glass plate inclined at 45° to the incident beam (which is perpendicular to the optic axis of the system in the manner usually adopted for showing Newton's rings by reflexion) then the intensity of the fringe pattern is some thirty times as great as that of the transmitted system. A narrow range colour filter can safely be employed without undue increase in exposure. The photographic times for the reflected fringes illustrated here were of the order of only 2 sec.

Reflected contour patterns are shown in figures 5 and 6 (plate 1). The contrast is not as good as with the transmitted fringes but the definition is decidedly better, when attention is given to fine detail. The reason is that the collimation can be made more perfect for the reflected system. Internal defects in the body of the crystal, as well as surface defects at the incidence surface, both contribute to reducing the

efficiency of collimation for the transmitted system. The reflected fringes are remarkably sharp, and in fact sharper than the corresponding Fabry-Perot fringes given by optical flats with the same silver coating. (The reason is that the wedge fringes are directly localized whilst the Fabry-Perot fringes are integrated over the whole interferometer surface and inevitable defects increase the final observable fringe width. Furthermore, the parallelism in the Fabry-Perot interferometer can never be perfect.) It might therefore appear that the reflected fringes should always be used in preference to the transmitted system. Actually the observation of the transmitted fringes is much easier; the photography of 'absorption fringes' is also rather a critical matter from the viewpoint of correct exposure; and in many cases (e.g. mica and selenite, see part II) it is necessary to employ more than one wave-length to decide the orders of fringe displacements. Thus the transmitted system, despite a slight inferiority in definition, offers advantages in particular problems. In fact both systems have their place in the experimental procedure, according to the requirements.

The reflected fringes in figures 5 and 6 complete the description of the (100) surface by extending the pattern right up to the edges of the face.

The perfection of the reflected system, with the wealth of significant detail revealed, shows that the experimental procedure described here can be very widely applied to opaque crystals, to badly transmitting crystals, to metallic surfaces, etc. and thus opens up a considerable field for investigation. It will be shown in what follows that not only are the major contour features of significance, but in addition all the small irregularities, kinks, angle bends, distortions, etc., are real, and reveal details of the surface topography.

THE INTERPRETATION OF THE FRINGE PATTERNS

(a) *Major features*

Comparison of figure 1 and the plates shows the interesting fact that *all* the markings in figure 1, just observable under a lens, correlate with detail in the fringes. The interference pattern goes, of course very much further, giving considerable additional information about the markings. The principal contour details will now be briefly reviewed. The lettering of figure 2 will be used throughout. To the left runs a long sharp ridge *c-d* about 2 waves in height. This terminates at the base in a triangular facet *B*. These features are responsible for the large triangular contours in figures 4 and 6. The facet *B* has a sub-facet at the right base which is very small. The sloping sides of the main ridge, *A* and *C*, are not flat, but are slightly convexly curved, as shown particularly by the contours in figures 3 and 5. The radius of curvatures of the sides is of the order of 60 m.

To the right of the ridge is the long strip *D*, of roughly rectangular shape. This is not flat but is convex, with a radius of curvature of some 33 m. The lower part of *D* is followed by a sloping drop *E*, then this rises to *F* in a sweeping upward rise. This surface *F* is curved on the upper half, being cylindrically convex with a radius

of curvature of some 21 m. The lower half is plane, and slopes down below the level of the rest of the surface towards the direction of the front edge of the whole face. The whole contour is encompassed within a depth of some four waves, i.e. $1/500$ mm.

The angles between a number of the larger faces have been measured and are given in the following table. The notation of the faces and ridges is that given in figure 2. Two ridges, *cd* and *ac*, show variation of face angle along the ridge length and this is shown separately. The angles given, in *minutes of arc*, are the angles between the normals of the respective faces.

TABLE 1

<i>Angles between major topographical features</i>												
face angle min. of arc	<i>CD</i>		<i>FD</i>		<i>DE</i>		<i>EF</i>		<i>AB</i>			
	0.65		0.65		0.75		0.50		2.4			
<i>Face angle AC. Variation of face angle along ridge cd.</i>												
<i>Distances measured from the point c</i>												
face angle min.	9.0	4.6	3.1	2.8	2.5	2.2	2.1	2.2	2.4	2.2	2.0	1.5
distance mm.	0	0.6	1.0	1.6	2.1	2.8	3.5	4.0	4.6	5.2	5.9	6.6
<i>Face angle CB. Variation of face angle along ridge ac.</i>												
<i>Distances measured from the point c</i>												
face angle min.	8.3				6.4				3.0			
distance mm.	0				0.7				1.4			

A number of the angles are *less than a minute of arc*. It is of interest to note that there is a marked variation along the straight ridges *cd* and *ac*. The sides of the ridge *cd* are also curved, as indicated previously. The face angles rise up in the vicinity of the apex *c*. The tetrahedral pyramid of which *c* is the apex has therefore concave sides in the neighbourhood of the point peak *c*. The face angles associated with this pyramidal feature are greater than those associated with any other major surface feature.

It is of interest to point out the extraordinary sensitivity of this procedure for the determination of a small angle. On a typical plate the angle bend between *fringes* for the edge *CD*, is 40° . The face angle being 0.65 min. of arc, it follows that a *fringe angle bend* of 1° corresponds to a face angle as small as 0.016 min. of arc. It should be noted too that these very small angles can be evaluated over quite moderately small areas of the surface. Comparison with the Rayleigh limit is of interest. For a face 1 mm. wide the smallest angle that can be measured by simple reflexion (Rayleigh limit) is 2 min. of arc, and for narrower faces the angle is proportionately greater. It will be noticed that 1 mm. corresponds to about one-tenth of the whole pattern registered here. No goniometer could possibly show up such fine detail as reported here, in particular the detection of such fine detail features as variations of angle *along a ridge*. The superiority of these measurements over those obtainable with the goniometer is of course due to the employment of



FIGURE 3. Contour fringes of quartz (100) face transmission.



FIGURE 4. Contour fringes of quartz (100) face transmission.



FIGURE 5. Quartz (100) face reflected fringes.



FIGURE 6. Quartz (100) face reflected fringes.



FIGURE 6. Mica (*a*). Contour fringes.



FIGURE 7. Mica (*b*). Contour fringes.



FIGURE 8. Mica. Detail of fringes of sample (*b*).

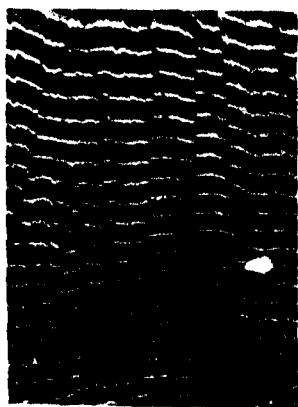


FIGURE 9. Selenite. Contour fringes.



FIGURE 10. Selenite. Detail shown by higher dispersion.

multiple beams. The Rayleigh criterion for resolution applies to a \sin^2 intensity distribution and it is the employment of a much more advantageous intensity distribution which permits such a higher degree of resolution to be achieved.

It seems clear that these *major features must be regarded as fine-scale vicinal faces*.

It is to be noted that *plane faces are the exception rather than the rule*. Despite the apparent considerable scale of the major contour features, when the enormous magnification is recalled it will be recognized that in effect the crystal face is really fairly flat.

(b) *Minor features*

Comparison between figure 1 and the various fringe patterns reveals novel facts concerning the minor markings on the surface. These markings are mainly of three broad types, namely (i) simple almost parallel striations, (ii) angle bend marks of the form \wedge , which are a characteristic feature of the (100) face, and (iii) features closely similar to (ii) but with the base closed to form an obtuse triangle so \triangle .

(i) *Striations*

The striae are only well marked on the sides of the major ridge $A-C$. Some are much fainter than others. Where the fringes pass over the more clearly marked striations a small kink appears. These kinks are very small indeed, some being clearly seen in the extended narrow triangle on the left of figure 5. Each observable kink is a significant detail, correlating with a visual striation mark. From the direction of the kink (whether the acute angle is in the direction of increasing or diminishing order of interference), it can be decided whether the striation represents a tiny rut or a small ridge. The pair in the middle of the triangle are ridges, the pair near the apex are ruts. In both cases, whether ridges or ruts, the maximum difference (i.e. the top of the ridge or the valley bottom of the rut) from the plane level whereon the feature occurs is extremely small. So small are the depths or heights, that only on the reflected fringes showing the best definition can accurate determinations be made. Measurement shows that the depths (heights) are only between 0.04 and 0.05 of an order, i.e. only some 100 Å units. *This corresponds to only 20 silica molecules.*

(ii) *The \wedge shaped markings*

Little information could be directly derived about these markings. A discontinuity in level takes place in their neighbourhood. It is a difficult matter to place fringes in the precise positions necessary to contour these small markings as separate units. It can only be concluded that in the near neighbourhood of the marks there are abrupt small changes in level, of a small fraction of a wave.

(iii) *The triangular markings*

Facts of some considerable interest were obtained about these. Complete small triangular fringe patterns are visible in the lower right-hand of figure 4 and more obviously on figure 6. It follows that the small triangular markings are *small*

tetrahedral pyramids with their bases resting on the surrounding main surface. These tetrahedra are *projections, not pits*. Measurements made on three such clearly defined markings give some 0.17 of an order as the approximate height of a pyramid, i.e. some 450 Å units. As the size of the pyramid base is the relatively large, easily seen triangular marking, it follows that the faces are inclined at very small angles to each other.

A significant point is that the small triangles on figure 6 closely resemble the major triangles arising from the three surfaces *A, B, C*, to the left of the same diagram. The face angles on the small pyramids are therefore very close to those enclosed between the sides *ABC*, i.e. between 2 and 3 min. of arc. As with the major feature, the pyramids can be regarded as a ridge meeting a triangular facet, or alternatively the major ridge feature *ABC* can be regarded as a tetrahedral pyramid. It is reasonable to conclude that both features are identical but on very different scales. It is probable that the small pyramidal projections represent the nuclei of growth of the (100) face, which have of course been arrested. The actual growth of layers from such nuclei has been recently directly observed in many crystals by C. W. Bunn (private communication).

DISCUSSION

That the visual markings have been proved to have depths (heights) which are only a small fraction of a wave-length is a matter of some interest optically. The fact that they are so easily seen requires explanation. It cannot be assumed that the markings appear *only* because of a differential reflectivity between the marked area and the surrounding surface. This is shown by the fact that the various markings can still be clearly seen after silver has been deposited on the crystal face. The reflecting coefficient is then uniform, i.e. that of the silver. Proof of this lies in the fact that the fringes retain their sharpness whilst passing over the relatively large marked features (see for example the small triangles already discussed). It can be concluded that a differential surface *quality* enables the markings to be seen. If it be assumed that the surfaces within the areas of the markings are ribbed or striated, either microscopically or even sub-microscopically, then in a sense they will behave as crude gratings, whilst the surrounding lustrous areas act as plane mirrors. Two facts support this explanation. The angle of illumination required to render the markings clearly visible is somewhat critical and small changes in the angles make the markings appear to change over from bright to dark, a typical grating characteristic. Furthermore, the markings are more clearly seen when the crystal surface is viewed through a colour filter. It is only necessary to postulate the association of a high lustre with the true (100) face to account for the difference in character of the faces of the small tetrahedral pyramids, since the latter are vicinal faces.

It has been suggested to the writer that some of the observed markings can be seen because of the light diffracted from their edges, in accordance with Rayleigh's

views on the visibility of discontinuities. This is a very likely explanation of the visibility of the ridge edges (junctions of vicinal faces) and of the striated markings. It is not likely to be the direct explanation of the visibility of the small tetrahedral pyramids since these extend over an appreciable area and it is quite clear visually that the whole of the area, and not only the boundaries, have a marked non-lustrous character. If the pyramids consist of stepped ridges, and this is very probably the case, then in a sense the Rayleigh mechanism can be considered to operate from each sub-microscopic step. In this case the final effect is not very different from that of the pseudo grating considered above.

I wish to express my thanks to Mr S. O. Agrell of the Geology Department, Manchester University, who kindly lent the quartz crystal.

The topography of crystal faces

II. The topography of cleavage faces of mica and selenite

BY S. TOLANSKY, *University of Manchester*

'Communicated by Sir Lawrence Bragg, F.R.S.—Received 14 April 1944

[Plate 2]

The multiple beam interference procedure described in part I is applied to the examination of the topography of the cleavage surfaces of mica and selenite. The precision attained is extremely high, particularly in the case of mica.

Mica. All the mica specimens examined exhibit considerable hills and dales which may extend up to heights and depths of $1/200$ th mm. in extreme cases. Sharp 'cleavage lines' are revealed, representing discontinuities in level (steps). These vary in length from quite small values up to several millimetres. The smallest step measured is 41 Å, which is only 2 'molecules'. Evidence is given showing that the steps are integral multiples of the 'molecular' thickness, 20 Å. Steps varying up to some 577 molecules are recorded, the majority being under 20 'molecules' in height. The error in the evaluation of the step, in the best cases is only 3 Å units. Up and down steps occur in a random way.

The value of a step remains constant along the length of a 'cleavage line', proving that the surfaces separated by this line are 'parallel'. The area of surface between a pair of cleavage lines is highly uniform and is at least true to 30 Å, being probably true to a molecular plane. The observations reveal objectively the 'invisible steps' found on mica by growing crystals on it.

Selenite. A selenite cleavage surface reveals a large number of long 'cleavage lines' which are mostly roughly parallel, although some are inclined at angles up to 15° . The characteristic hills and dales shown by mica are absent, but the specimen investigated exhibited a cylindrical curvature, being curved in the same direction as the 'cleavage lines'. The observed steps at these lines vary from 16 to some 146 molecules in height. The faces on the adjacent sides of a line (in contrast with mica) are usually inclined to each other, the slope being in the direction of the line. The angles of slope are minute, the measured values varying from 0.011 to 0.063 min. of arc.

The fringes are very ragged and are nowhere smooth as in mica. High dispersion shows that this is due to the existence of a secondary structure consisting of a very large number of small elongated facets. The long dimension (at least ten times the width in most cases) is in the direction of the cleavage lines. These facets are mostly of the order of $1/50$ th mm. wide and vary in level in a random way by only a few 'molecules'.

INTRODUCTION

It has been shown in part I, which describes the interferometric examination of the topography of a (100) face of a quartz crystal, how multiple beam interferometry can be used for the examination of the topographical features of crystal faces which are reasonably plane. In this paper the technique described there is applied to the examination of the cleavage faces of mica and of selenite. The present work is but a preliminary survey, and no attempt is made to correlate the very complex features observed with crystallographic data. It is hoped that such a more detailed correlation will be undertaken in the future. General information of considerable interest has been derived and it will be clear from what follows that the interferometric approach should throw light upon the mechanism of cleavage.

Cleavage planes of both mica and of selenite are well suited to the transmission fringe technique described in part I. Such excellent collimation can be achieved with thin slips of these crystals that very high grade definition results, and, as it is necessary to employ more than one wave-length for the identification of orders, only the transmitted system has been employed.

With both crystals the treatment of the surface was the same. An *untouched* freshly cleaved surface was coated with a high-reflecting-coefficient silvering by evaporation *in vacuo*. No mechanical cleaning was used, the only cleaning attempted being ionization de-gassing by a glow discharge. This improves the quality of the silver surface which is deposited. The silvered crystal faces were matched against the optical flat that was used for the examination of quartz.

TOPOGRAPHICAL FEATURES OF MICA CLEAVAGE FACES

A number of different samples of mica were examined. They were of unknown origin or classification (more comprehensive work will be undertaken later, with classified specimens). When thin sheets of mica were mounted close to the optical flat they were found to behave as highly sensitive diaphragms and were extremely susceptible to small air shock-waves. The fringes were never quite still. This was particularly noticed in regions of high dispersion. (It is clear that so sensitive a diaphragm, which indicates displacements of the order of $\lambda/100$, might have some useful applications.) This difficulty was easily overcome by lightly pressing the mica against the flat with a sheet of glass. The separation between the silvered surfaces is then very small and the fringe definition is thus excellent. There was complete stability over long periods, but there was loss in control in adjusting the tilt of the mica surface relative to the flat. Sufficient control was available by local light pressure on the glass backing.

Two characteristic mica contours, given by different samples, are shown in figures 6 and 7 (plate 2). The area represented is roughly a square centimetre. The stronger fringes belong to the green mercury line, the weaker doublet being the yellow lines. To assist identification, photographs were taken without any filters. Despite the obvious complexity of the patterns three characteristic salient features are evident, and such features appeared on all the mica samples examined.

(i) The general nature of the surface contour is, in all cases, similar. There are hills and dales (in some cases quite small pits) which sweep across the surface. Different samples vary only in degree, but not in character. Figure 7 shows, for example, numerous hills, some $1/200$ th mm. in height, whereas the sample of figure 6 is much more nearly plane.

There is some indication that, with thin mica slips, the general shape of the hill features is changed by the act of peeling off sheets behind the silvered surface. This is under examination, and if substantiated it would indicate that the hills and dales are linked up with the creation or release of internal strains.

(ii) A striking characteristic is the existence of sharp clear cut *cleavage lines* representing definite discontinuities in level. These vary considerably in length. The 'cleavage lines' are often fairly straight. In some samples they run in one general direction, in others they intersect, at times with large angles, up to almost 90° . It is obvious that an attempt should be made to correlate the directions of these lines with crystallographic data, and this is being undertaken.

The detail of a section of the sample shown on figure 7 is shown enlarged in figure 8. This shows both the green line (over-exposed) and the yellow doublet. The exposure is correct for the latter, which bring out the remarkable sharpness of the fringes. (The distance from one doublet to the next is an order, equivalent to half a wave change in height on the crystal face.)

(iii) There exist large areas between the 'cleavage lines' which are of *great optical uniformity*. One must distinguish between (a) distortion into hills and dales, and (b) local irregularities. The smooth run of the fringes between the 'cleavage lines' shows that, although the surface may be locally badly distorted into a sharp curve, it remains smooth and uniform. Of particular interest is the manner in which some 'cleavage lines' pass straight over hill and dale features. This would seem to indicate that the cleavage *first runs along a surface, which subsequently distorts because of the release or creation of strain*.

THE CLEAVAGE LINE DISCONTINUITIES IN MICA

Each cleavage line represents a sharp discontinuity in level. Whether the drop (or rise) is at right angles to the surface or slopes, it is impossible to decide, for the lateral magnification is negligible. By comparing the fringe patterns given by lines of different wave-length, the exact displacement (whole orders plus fractions) can be obtained without any ambiguity. The use of more than one wave-length is essential too, to decide the *direction* of the step, i.e. whether up or down.

In the large majority of steps the displacement is less than a whole order, that is, the two surfaces differ in height by less than half a wave. It is of particular interest to evaluate what is the smallest step that can be measured on the sharpest fringes. The smallest step detected and measured on plate 1 is 0.015 of an order, *which is only 41 Å units*. It is estimated that a *step* of 30 Å units could just be detected under the best of conditions.

There is no question about the reality of these small steps. A step cannot be confused with possible local errors on the flat since it extends in a straight line across a number of fringes, and the value of the step derived from any pair of fringes is constant.

It is to be noted that, because the stepped fringes are displaced above as well as sideways, the resolution far exceeds that of corresponding Fabry-Perot fringes, in which ring displacements are directly between the orders. The situation is analogous to that which obtains in the use of fiduciary fringes such as are used in the Rayleigh refractometer.

In order to give a survey of the form taken by the cleavage discontinuities, the contour of a traverse across selected local portions of the micas shown in figures 6 and 7 are shown respectively in figures 1 and 2. Tables 1 and 2 give the measurements. The figures represent a traverse of several millimetres on the crystal face in a direction approximately at right angles to a group of cleavage lines.

The fraction of an order between displaced fringes was evaluated by the standard McNair procedure employed in hyperfine structure analysis. In approximately plane regions, where the dispersion is almost linear, extremely high precision can be attained, an accuracy of one thousandth of an order being obtained without difficulty.

Owing to the absence of flexibility in the mounting adopted to overcome vibrations, there is no easy way of deciding rapidly whether the diagrams in figures 1 and 2 are the right way up or whether they should be turned upside down. The steps on the two mutually cleaved faces should clearly be complementary, and if this is true the direction of the figure is of no consequence, for one pattern will refer to one mica sheet and the upside down pattern to the other. This question will be checked in future work. (*Note added 1 June 1944. A method for deciding this point has now been developed and will later be described.*)

It is clear that, unless consideration be restricted to steps measured with the highest precision, it will be easy to arrive at a false picture of steps occurring in integral multiples of some unit. Conclusions can only be drawn about the ratios of steps from those measured with precision. The factor which affects precision is mainly the influence of the hills and dales affecting the *dispersion locally* and making it difficult to reduce an accurately measured displacement into an accurate fraction of an order. There are five steps in which the precision of measurement is very great. These will be discussed later.

An important feature concerns the value of a step measured successively from a run of fringes *along a cleavage line*. If measurements are made in those regions

not badly distorted by hills and dales, it is found that the fringe displacement along the length of the cleavage line is uniform, indicating that the step is constant so that *the two faces separated by the cleavage line are parallel, or, if curved, have the same curvature.*

The general hill and dale contour of the whole surface sets a limit to the accuracy with which this can be determined, but it appears that the parallelism is fairly exact locally, at least over a number of adjacent fringes, and certainly to within a very small fraction of an order.

THE LOCAL UNIFORMITY BETWEEN THE CLEAVAGE LINES OF MICA AND THE STEPS

The *local* uniformity between cleavage lines is a feature of some interest. The following can be deduced from figures 6 and 7. The green-line fringes are over-exposed to bring out the yellow mercury doublet. The doublet separation is less than 1/10th of an order. The local 'wiggles' in each fringe are of the order of 1/10th of the doublet separation (perhaps slightly more). Thus local fluctuations are of the order of about 1/100th of a fringe, at most 1/75th. This amounts to between 1/200th and 1/150th of a wave and is clearly the limit set by the optical flat. Thus it can be concluded that, disregarding general superposed curvature, the mica surface must be uniform to be better than 40 Å, and probably than 30 Å.

In a private communication, Sir Lawrence Bragg has pointed out that the cleavage lines and steps described here are almost certainly to be identified with the 'invisible steps' inferred to exist on some mica surfaces from experiments made by Friedel on the growth of ammonium iodide crystals upon such surfaces (see Sir Lawrence Bragg, *Atomic Structure of Minerals*, p. 210, 1937). When these crystals are grown on mica surfaces, with some samples the crystals orient themselves oppositely along invisible dividing lines. It has been concluded from these experiments that over the area of identical orientation the mica surface has cleaved true to within a single molecular plane. According to Sir Lawrence Bragg, as the mica cleaves along the plane in which the K atoms are situated, and as the height of the molecule (strictly, the *c* spacing) is 20 Å, it follows that all the cleavage steps should be integral multiples of 10 Å, since the K atoms are centrally situated.

The interference data can now be compared with these conclusions. The smallest step recorded is 41 Å, i.e. exactly twice the *c* spacing (effectively two 'molecules'). There may be smaller steps, but these are on or beyond the limit of measurement. At the other end, the upper limit of step has not been evaluated since the interference procedure is unsuited to the measurement of the larger steps which come within the range of the microscope. A step exceeding 9000 Å (about 1/1000th millimetre) is recorded in figure 1.

In tables 1 and 2 the complete range of steps observed is given in the last column in terms of 'molecules' (multiples of the *c* spacing which is taken to be 20 Å, a value sufficiently exact for the present purpose, see *Atomic Structure of Minerals*). The observed steps range from 2 to some 577 molecules.

There are in these tables, in all, 27 measured steps. Amongst them are 5 in which the precision of measurement is extremely high since they occur in areas nearly plane. The fractional order displacement is known to 0.001, which corresponds to an error of the extremely small value of 3 Å units. It is a striking fact that *these*

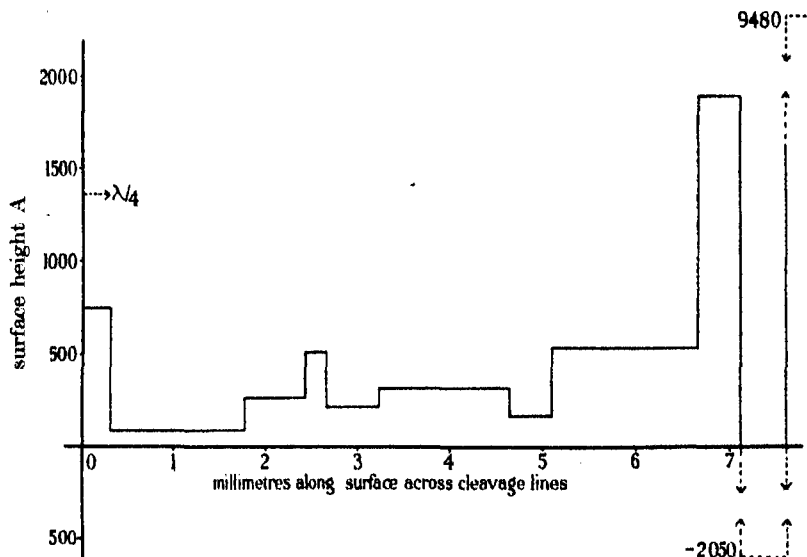


FIGURE 1. Mica (a), cleavage step contour.

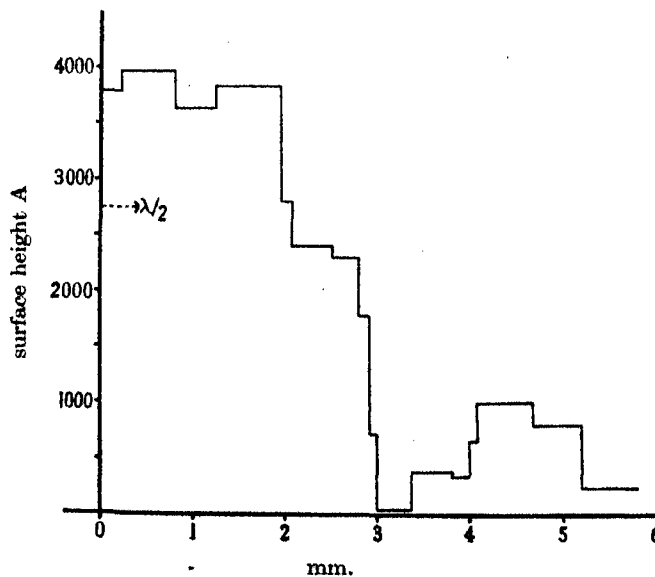


FIGURE 2. Mica (b), cleavage step contour.

steps are exact integral multiples of 20 Å (within the error). These steps are, in angstrom units

41	100	158	180	341
20×2.05	20×5.00	20×7.90	20×9.00	20×17.05

Furthermore, these steps all being *small* multiples of 20 Å, there is no question about the certainty of the integral ratios. *It is certain then that the steps are simple multiples of whole 'molecules'.* There is, at least amongst these 5, *no evidence of 10 Å being the fundamental unit.* This might only be a matter of chance, for clearly if 10 Å happens to be the unit, it is simply a question of whether amongst 5 random steps, all will have an even number of units. The number of steps is too small to draw a safe conclusion, especially when it is noticed that 3 are from one sample and 2 from the other sample, but the evidence seems to incline to favour 20 Å as the unit, in these small areas.

TABLE 1. MICA (a)

position on crystal mm.	step, fraction of order	step, angstrom units	step, molecules, as whole numbers
0	$+0.183 \pm 10$	499 ± 27	~ 25
0.32	-0.203 ± 2	553 ± 5	28
1.80	$+0.063 \pm 2$	172 ± 5	9
2.45	$+0.087 \pm 3$	238 ± 8	12
2.67	-0.107 ± 3	292 ± 8	15
3.25	$+0.0369 \pm 5$	100 ± 1	5
4.67	-0.057 ± 1	156 ± 3	8
5.12	$+0.132 \pm 2$	360 ± 5	18
6.65	$+0.246 \pm 2$	671 ± 5	34
7.12	-1.188 ± 2	$3,242 \pm 5$	162
8.02	$+4.23 \pm 1$	$11,535 \pm 27$	~ 577

TABLE 2. MICA (b)

position on crystal mm.	step, fraction of order	step, angstrom units	step, molecules, as whole numbers
0	-0.066 ± 1	180 ± 3	9
0.25	$+0.118 \pm 5$	322 ± 13	~ 16
0.84	-0.066 ± 2	180 ± 5	9
1.27	$+0.376 \pm 5$	1026 ± 13	51
1.97	$+0.144 \pm 20$	393 ± 50	~ 20
2.09	$+0.041 \pm 3$	112 ± 8	6
2.54	$+0.197 \pm 4$	538 ± 10	~ 27
2.84	$+0.380 \pm 20$	1040 ± 50	~ 52
2.97	$+0.248 \pm 4$	677 ± 10	~ 34
3.00	-0.125 ± 1	341 ± 3	17
3.40	$+0.015 \pm 1$	41 ± 3	2
3.81	-0.115 ± 3	314 ± 8	16
4.00	-0.127 ± 4	347 ± 10	~ 17
4.10	$+0.080 \pm 2$	215 ± 5	11
4.72	$+0.195 \pm 2$	532 ± 5	27
5.22	$+0.142 \pm 2$	387 ± 5	19

It should be noted that this in no way contradicts the ammonium iodide crystal growth experiments. The successive double layers of mica are related by a glide plane of symmetry parallel to (010). The orientation on the surface for any step may have an inclination on either side of this symmetry plane. On adjoining steps the orientations may be the same, and then the crystals of ammonium iodide would all point the same way. The step would then be an integral multiple of 20 Å, and, if this happens, the ammonium iodide experiment would in fact *not reveal a step*. Three illustrations of crystal growth are given by Sir Lawrence Bragg. One, lepidolite, shows a clear step, hence this is a half integral multiple of 20 Å. But another sample shows no 'invisible' step. It does not follow that the surface is uniform, it may in fact have on it a number of steps, each integral multiples of 20 Å. The mica samples used here could give ammonium iodide patterns similar to this second sample. Finally the third crystal growth illustration (phlogopite) shows orientations of either kind equally numerous, from which it might be concluded that here we have a typical area in which there are many small length cleavage steps, of narrow width between steps. Such characteristic areas have appeared on some of the samples examined here.

Not only is there agreement between the general conclusions as regards steps, obtained by interferometry and by crystal growth experiments, the same is true of surface uniformity. The existence of uniformity between the 'invisible cleavage' lines given by the crystal growth orientation is not quite a safe inference, for, as already pointed out, there exist steps which are integral multiples of 20 Å and from the point of view of crystal growth such steps would have no effect at all on orientation and thus be missed, giving a false picture of uniformity.

It was concluded from the smoothness of the doublet yellow fringes that, between cleavage lines, the surface is, however, actually uniform to at least 30 Å. Thus the uniformity can be objectively *demonstrated* with fringes although only deduced from the crystal orientation (although the orientation observations are subject to the doubt already mentioned). It seems that the conclusion drawn from the crystal growth, namely that between the cleavage lines the surface runs true to a molecular plane, is substantiated. It is clearly more logical to assume this than to postulate the possibility of irregular steps of either only one or of only two molecules, indiscriminately, which is the limit set optically by the fringes.

There is definite evidence that on specimens having high hills of the order of 10 waves, some hillsides do not fall smoothly, but ripple. This is shown by the splitting of fringes into close multiplets. This feature will be examined further. It can be clearly seen on figure 7 and 8 (plate 2).

Finally it may be noted that if the mica sheet is reversed so that there is a thin slip of mica between the silvered surfaces, then double fringes, due to the double refraction of the mica, become visible. Alternatively, if both sides of a mica sheet are silvered and the optical flat dispensed with, then complexity results because, in addition to the doubling of the fringes due to double refraction, the cleavage steps on the two separate surfaces come into play and cannot be disentangled.

without separate observations made on each face, against a flat. This procedure reveals important internal and surface features, however; and is now being developed in detail.

THE TOPOGRAPHICAL FEATURES OF SELENITE CLEAVAGE FACES

The specimen of selenite available gave fairly good specular reflexion, before silvering, over some 2 sq. cm. A typical contour, representing about 1 sq. cm., is shown in figure 9 (plate 2). The contrast with mica is striking. The main distinguishing features of the pattern are:

(1) A large number of clearly marked cleavage lines, which are roughly parallel. The large angles exhibited by some micas do not appear, although in isolated cases angles up to some 15° appear. It is hoped to correlate these directions with crystallographic data, if this is possible.

(2) The characteristic hillocks and valleys shown by all mica specimens are absent. Instead, the selenite shows a cylindrical curvature. The curvature is relatively large, the radius being some 4.6 m., for one particular sample. *The cleavage lines are in the same direction as the curvature.*

(3) The separate fringes are very ragged and show multiple kinks. Between the straight cleavage lines the fringes are *nowhere* smooth as was the case with mica but are disjointed and irregular, although retaining their individualities as separate fringes which are still sharp.

The detail of the separate fringe structure is shown in figure 10 (plate 2). This is not only a slightly bigger enlargement but the dispersion has been considerably increased by bringing the crystal surface nearer to parallelism with the optical flat.

THE CLEAVAGE LINE DISCONTINUITIES IN SELENITE

The ragged nature of the fringes reduces the high precision of measurement when comparison is made with the data for mica, yet the precision is still considerable. The contour of a typical traverse across 7 mm. of the crystal, at right angles to the cleavage lines, is shown in figure 3, the measurements being given in table 3. The numbers of 'molecules' in the steps are approximate whole numbers, for the accuracy is insufficient to give an exact figure.

There exists a close superficial resemblance with the typical mica contours of figures 1 and 2. A very large number of steps smaller than those recorded above were observed, but they were not measured since they extended for only short lengths and tended to become merged into the general ragged nature of the fringes. It will be noted that the dimension of the selenite cell in the direction perpendicular to cleavage is 15 Å, which compares with the 20 Å of mica.

VARIATION OF STEP ALONG A SELENITE CLEAVAGE LINE

A very marked difference between mica and selenite is that the value of the step, in many cleavage lines, changes regularly on moving *along* a selenite cleavage line. This means that the two surfaces separated by the cleavage line are *not*

parallel, having a relative slope *along* the direction of the cleavage line. A particularly marked case is shown by the data given in figure 4. This is the plot of the value of the cleavage step *along* a 3 mm. length of a cleavage line. This shows that, (a) there is a relative small curvature between the two faces, (b) this is superposed on a general slope. Taking one surface as the horizontal axis, the other starts off below this, crosses over, and a downward step becomes an upward step.

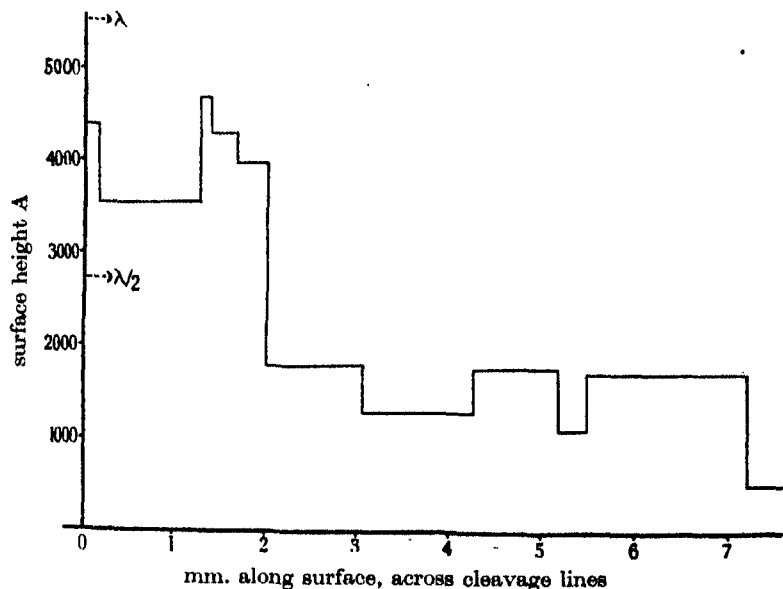


FIGURE 3. Selenite, cleavage step contour.

TABLE 3. SELENITE

position on crystal mm.	step, fraction of order	step, angstrom units	step, molecules, as whole numbers	slope of adjacent faces, min. of arc
0	+0.307	838	56	-0.061
0.16	-0.414	1130	76	—
1.26	+0.143	391	26	+0.017
1.38	+0.112	307	20	-0.011
1.78	+0.802	2188	146	—
2.01	+0.184	501	33	-0.018
3.06	-0.176	480	32	—
4.27	+0.265	674	45	-0.018
5.18	-0.225	615	41	-0.063
5.48	+0.090	244	16	—
7.22	-0.351	957	64	+0.017

The case illustrated in figure 4 is an extreme example for in the majority of the cleavage steps measured any relative curvature that may exist between the adjacent faces is so small, that it is sufficiently close to consider that the two surfaces are merely sloped relative to each other in a linear manner.

The numerical values of the slopes, in *minutes of arc*, are given in the last column of table 3.

The character of the sloping surfaces is illustrated by an approximately perspective diagram, figure 5. This represents the same contour as shown in figure 4,

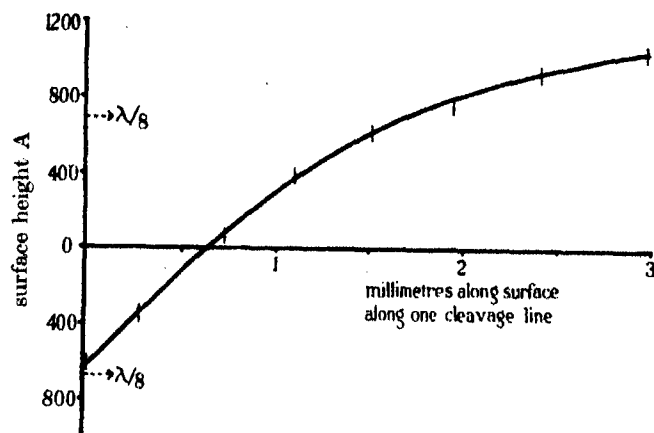


FIGURE 4. Selenite, variation of step along cleavage line.

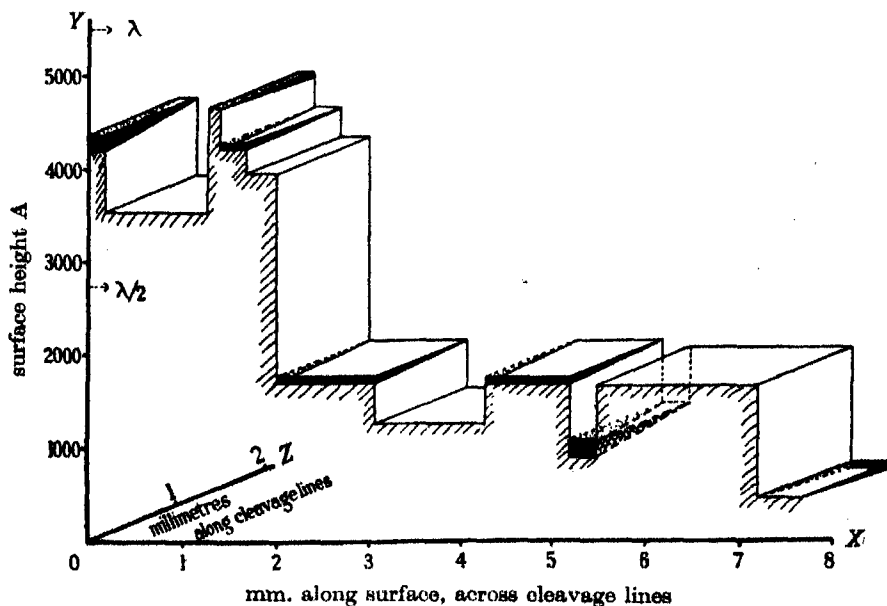


FIGURE 5. Selenite, cleavage step contour.

with the third dimension added. The axes OX , OY , OZ , are mutually perpendicular. The scales along the two horizontal directions are the same, the magnification only being great in the vertical direction. The deviations of the surfaces from the horizontal plane are shown by the blackened wedge sections. The thickness of the

slab along the *OZ* direction is 1 mm., hence the diagram represents the variations along the cleavage lines of a length of 1 mm. *There are both positive and negative slopes* (positive direction arbitrary) and on some planes the slope is too small to be determined with certainty. It is of interest to note that the two surfaces which occur at *X* approximately 2–3 mm., and at *X* approximately 3–4 mm., have, as near as can be estimated, both the same level on the *Y* axis and the same slope. It is clear that the portion between them has been simply torn out of the area extending over *X* approximately 2–5 mm.

The angles of the sloping faces are of course very small and it is only the very high magnification in the *Y* direction which enables them to be measured. It may be noted that the slope of only 0.011 minute of arc was given by a crystal strip 0.8 mm. long and 0.4 mm. wide.

THE STATE OF THE SURFACE BETWEEN THE SELENITE CLEAVAGE LINES

The peculiar state of the surface between the long cleavage lines is revealed by the higher dispersion of figure 10, especially in the region of considerable dispersion near to the centre of the cylindrical fringe pattern. This plate shows why, with lower dispersion, the fringes have so ragged an appearance. It can be clearly seen that the areas between the major cleavage lines consist of a large number of small facets, each of which is an elongated strip. The long axis of the strip is in the direction of the major cleavage lines. The strips vary in length from quite small strips to lengths up to 1 mm. The widths also vary widely and, although most are of the order of only 1/50th of a millimetre wide, some attain a width of the order of 1/10 mm. These strips *do not lie in the same plane* but are stepped in a haphazard way, positively and negatively. The steps are only very small fractions of an order in most cases. It is clear that the strips *differ in height by only a few molecules*.

There is no evidence of any regularity at all in this distribution of small strips or in their heights. Occasionally an individual narrow strip will project out of a fringe by as much as 1/10th of an order, indicating that there is a difference in height by some 300 Å, i.e. some 20 'molecules'. In general the differences are considerably less.

The individual small elongated strips practically all lie with their major axis in the general direction of the long cleavage lines, but in some cases there are appreciable abrupt angles between adjacent small strips. The general direction of the large majority proves that the strips *must be regarded as subsidiary cleavages superposed upon the main areas defined by the major cleavage lines*. The strip edges are very sharp, proving true discontinuity.

DISCUSSION

Selenite is a hydrate and the crystal has been exposed to a hard vacuum for some 30 min. during the silver deposition procedure. It must be considered first whether some slight surface dehydration effect might account for the ragged

surface. The crystal never warms up above room temperature during the silvering procedure so that the dehydration, if any, might be expected to be minute. On the other hand the facets are only a few molecules deep. It seems certain, however, that dehydration has nothing to do with the effect. The fact that the individual strips show such extremely sharp cleavage edges, and the fact that these are in the same general direction as the major cleavage lines, disposes of this view. The subsidiary strips must certainly arise during cleavage, indicating a true secondary structure.

Thus it is clear that the cleavage mechanism differs from that of mica in that there are two distinct factors (*a*) the major strong cleavage strips extending over distances of the order of centimetres, and (*b*) the secondary weak cleavage patches which are in most cases only a fraction of a millimetre in length. In mica the cleavage takes place at a weak bond between potassium (or sodium) and oxygen, whereas in selenite the weak bond is between water and oxygen. This difference may be connected with the different cleavages. On the other hand, whether or not the subsidiary patches in selenite are related to a crystal mosaic remains yet to be determined.

The contrast between the mica and selenite cleavages is further accentuated by the common occurrence of a relative slope between the two faces on either side of a cleavage line in the case of selenite. It is also fairly certain that in selenite the change in level as one moves *along* a cleavage line is not continuously smooth but rather stepwise and ragged, because the subsidiary patches are so much shorter than the main cleavage lines and they lie at variable heights.

It is clear now that the three dimensional diagram in figure 5 is only a simplified schematic representation of the true state of the surface.

I wish to express my thanks to Mr S. O. Agrell of the Geology Department, Manchester University, who kindly supplied me with the selenite used.

Note added (10 February 1943). After completion of this work, a preliminary short note giving a brief survey of the results obtained was communicated to *Nature* (152, 722, 18 Dec. 1943). As a consequence of this a communication was received from Professor Manne Siegbahn, of Stockholm, who sent the author a copy of a paper published by Siegbahn in *Ark. Mat. Astr. Fys.* 23 A, no. 12 (1933). This describes an interferometric procedure which was employed for the examination of crystal cleavage planes. The author was not aware of this publication since the Journal in question is not available in Manchester.

The experimental procedure described by Siegbahn is the classical *two-beam* interference method of Fizeau. Whilst this much less powerful method is capable of revealing the coarser surface structure, the resolving power, being severely limited by the unfavourable \sin^2 intensity distribution, is insufficient for revealing the subtle details of structure shown by *multiple beam* interference. As a close analogy it can be considered that the resolving power of this multiple beam method is to that of the two-beam procedure as is the resolving power of a 30 line grating to that of a two line grating. The accuracy of measurement is correspondingly affected.

The crystal structure of cholesteryl iodide

By C. H. CARLISLE AND D. CROWFOOT

Department of Mineralogy and Crystallography, Oxford

(Communicated by J. D. Bernal, F.R.S.—Received 19 June 1944)

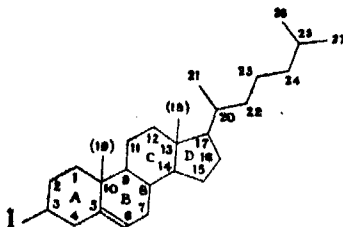
[Plate 3]

The X-ray analysis of cholesteryl iodide is an example of the use that can be made of the presence of a heavy atom, here iodine, in working out the structure of an organic molecule containing a number of asymmetric centres. The method of analysis employed has been to calculate the electron density within the unit cell, using the observed F values and, at first, phase angles derived from the contributions of the iodine atoms alone. The electron density pattern so deduced necessarily has a centre of symmetry, since the positions of the two iodine atoms in the unit cell are related to one another by a centre of symmetry. It therefore shows, somewhat approximately, the positions of all the atoms in the molecule superimposed on those of a spurious mirror image molecule. Selection of the atomic positions belonging to one molecule was made through a consideration of normal carbon-carbon bond lengths and valency angles. The analysis was greatly assisted by the comparatively simple crystallographic character of the compound. Cholesteryl iodide crystallizes in two forms, *A* and *B*, of closely related crystal structure, both monoclinic, $P2_1$, with $n=2$. Fourier projections on (010) were derived for both forms and show clearly the outlines of the sterol molecules and their general arrangement. The resolution of the atomic positions is best in the case of *B* and three-dimensional analysis was therefore attempted on this form only.

The crystal structure derived confirms the general view of sterol crystallography put forward by Bernal and also the present accepted chemical structure of the sterol skeleton. The accuracy with which the positions of the individual atoms are fixed is not great, but it appears sufficient to establish their mode of linking and to provide new evidence on their stereochemical relations.

The present X-ray analysis of cholesteryl iodide (I) is an attempt to determine the detailed structure of the sterol skeleton using as little chemical information as possible. It is the natural continuation of the crystallographic measurements on sterols begun by J. D. Bernal in 1932 and illustrates some of the general problems involved in the application of X-ray methods to the determination of the chemical structure of complex organic molecules. In the interval since Bernal's first measurements the actual chemical structure of the sterol skeleton has been very largely established, but there still remain problems in detail, particularly of stereochemistry, which need further investigation. There is a particular interest in finding how far information can be gained on these points, quite apart from the general interest of attempting to check the structure as a whole by X-ray methods. On the other hand, it has become clear from the crystallographic side that for such a purpose some approach to a direct X-ray analysis must be sought. This is chiefly because the usual methods of trial and error analysis become almost impossibly difficult where the molecular structure is both complicated and at least partly unknown. The point has been illustrated already in the sterol series, both during the survey of a variety

of sterol crystal structures (Bernal, Crowfoot & Fankuchen 1940) and by the more detailed examination of, for example, the choleic acids (Caglioti & Giacomello 1939).



In the analysis of cholesteryl iodide we are making use of the fact that the phases of the X-ray reflexions are largely controlled by the contributions of the heavy iodine atoms. Fourier series calculated by the usual equation, (1) (Bragg, W. H. 1915; Bragg, W. L. 1929), using the observed F' values and the heavy atom

$$\rho_{xyz} = \sum_{-\infty}^{+\infty} \sum_{-\infty}^{+\infty} \sum_{-\infty}^{+\infty} |F_{hkl}| \cos \{2\pi(hx + ky + lz) - \alpha_{hkl}\}, \quad (1)$$

phase angles (α_{hkl}), should therefore give directly an approximate picture of the electron density (ρ_{xyz}) within the crystal structure (cf. Robertson & Woodward 1940; Huse & Powell 1940; and also Cox & Jeffrey 1939). How closely this picture approaches to the truth depends upon the actual crystallographic conditions present. The most favourable circumstances are those found by Robertson & Woodward (1940) in the crystal structure of platinum phthalocyanine where the platinum atoms are at a crystallographic centre of symmetry and by their contributions exactly determine the phases. In the general case, of which cholesteryl iodide is one particular example, this is not true; here both the molecular and crystal structure are asymmetric, the heavy atoms are not in special positions, and the true phase angles must differ from those calculated on the positions of the heavy atoms only. The most serious difficulty is that, in general, the positions of the heavy atoms in the crystal will be related to one another by higher symmetry than are the molecules as a whole—in the present example by a centre of symmetry. The phases deduced from their contributions alone, in turn, introduce this symmetry into the calculated structure. Consequently, although peaks should appear in the derived Fourier syntheses at approximately the correct atomic positions, they will also occur at additional positions, related to the first by the extra symmetry element. At best the use of these phases deduced from the heavy atoms alone only limits the position of any particular atom in an asymmetric crystal structure to certain alternatives.

One further point should be mentioned. In the molecule of cholesteryl iodide, unlike that of platinum phthalocyanine, the atoms have a non-planar arrangement, and X-ray analysis in three dimensions is essential.

The preliminary examination of cholesteryl iodide carried out during the sterol survey showed that this crystal structure had certain favourable characteristics

which might make a direct and detailed analysis possible, in spite of these complications. In the crystal unit cell there were only two molecules, related by a twofold screw axis of symmetry. And the fact that this symmetry axis was parallel to the least refractive index of the crystal suggested that it was roughly normal to the sterol ring system. A projection on a plane normal to the axis might therefore be expected to show well-resolved atomic positions. And further, since this projection is crystallographically centro-symmetrical, its accuracy should not be affected by the symmetry introduced into the three-dimensional structure by the use of the iodine phase angles.

With these considerations in mind, the following procedure was suggested for the systematic analysis of the structure of cholesteryl iodide:

- (1) Measure the intensities of all (*hkl*) reflexions.
- (2) Deduce the positions of the heavy atoms from them by the calculation of Patterson series (Patterson 1935).
- (3) Calculate the phases from the heavy atom contributions alone, and using these and the observed *F* values form Fourier series to reveal the positions of the remaining atoms. This process is conveniently broken into two: (*a*) the calculation of a Fourier projection on the crystallographic *b* plane which should give as much resolution as possible of the atomic positions in two dimensions; (*b*) the calculation of three-dimensional line series through the peaks revealed in the first Fourier series in order to determine their third parameters (Harker 1936; Goodwin & Hardy 1938).
- (4) Distinguish the correct alternative among the possible positions of the atoms shown so far by considerations of, for example, bond distances and valency angles.

- (5) Recalculate phases and repeat Fourier summations to correct the structure deduced at (3) and to determine as closely as possible the true atomic positions.

At the outset of the analysis a complication occurred. A second metastable crystal-line modification of cholesteryl iodide, *B*, was found in the preparation, very similar in unit cell dimensions and optics to *A* but differing from this in the intensities of the X-ray reflexions. Since to a first approximation both appeared equally suitable, detailed examination was begun on the two crystal structures together. The first Fourier projection obtained at stage 3 (*a*) showed greater resolution of the atomic positions in structure *B* than in structure *A*, and the complete three-dimensional analysis was therefore continued on *B* alone.

In the account which follows the tables numbered in italics have been omitted for reasons of space. Copies are deposited with the Royal Society and they also occur, together with additional information on the calculations carried out, in a thesis by one of us (Carlisle 1943).

STRUCTURE DETERMINATION

The sample of cholesteryl iodide used was that prepared by Beynon, Heilbron & Spring (1936) by the action of hydrogen iodide on *i*-cholesteryl ethyl ether and *i*-cholesteryl benzyl ether.* The first X-ray measurements were carried out on this

preparation by I. Fankuchen in Cambridge in 1937-8. At the start of the present investigation the crystals had become reddish brown in colour, presumably through partial decomposition, and the material was therefore recrystallized by slow cooling from hot acetone. X-ray examination showed that the recrystallized preparation contained two crystalline varieties, *A* and *B*, the proportion of which varied in different recrystallizations, suggesting that they were polymorphic modifications. Further evidence on this point was obtained by microscopic observations on the effect of heat on the two (hot-wire method). *A* transformed on heating to a metastable form, melting simultaneously with *B*. On cooling, the two preparations behaved exactly similarly, the metastable form grew in both melts followed by the stable form. By rapid cooling a preparation could be obtained showing the two forms in contact with one another for an indefinite length of time.

The preliminary X-ray data on the two forms are summarized in table 1. Once their difference was recognized it was found possible to distinguish them morphologically by inspection under the microscope. Their characteristics are as follows:

Cholesteryl iodide A. The crystals are small, almost colourless, monoclinic laths, elongated along $[010]$, showing $\{001\}$ dominating and $\{100\}$. The birefringence is positive, $2V = 56^\circ$ (in glycerine), α is parallel to b , γ lies approximately 6° to c in the obtuse angle.

TABLE 1

	a	b	c	β	ρ	space group	n
cholesteryl iodide <i>A</i>	10.93	10.34	21.47	149°	1.300	$P2_1$	2
cholesteryl iodide <i>B</i>	12.57	9.04	21.89	149°	1.275	$P2_1$	2

————— $\frac{1}{2}$ mm. —————

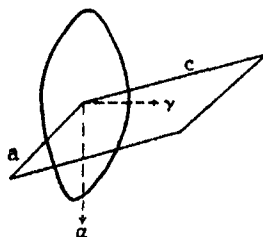


FIGURE 1. Cross-section of crystal of cholesteryl iodide *B* normal to the b axis.

Cholesteryl iodide B. The crystals are usually larger than those of form *A* and pale yellow in appearance. The colour deepens on keeping, probably owing to slight decomposition with liberation of iodine. They are irregular, platy, monoclinic crystals, elongated along $[010]$. Their usual cross-section normal to $[010]$ is as shown in figure 1, the forms developed being far from clear, probably $\{001\}$ and $\{10\bar{2}\}$. The birefringence is positive, $2E = 88^\circ$, α is parallel to b , γ lies in the obtuse angle at approximately 15° to c .

* Apparently the same compound has since been isolated by Helferich & Gunther (1939) by the action of HI on cholesteryl toluene sulphonate and cholesteryl methane sulphonate.

The density in both cases was measured by flotation in potassium mercuric iodide solution.

These preliminary observations suggested strongly that in both forms an open projection of the sterol ring system would be obtained on (010), normal to the symmetry axis. The analysis then followed the scheme suggested above. In both structures $F_{000} = 520$, $2f_0I = 106$.

(1) *The estimation of the intensities*

Two types of X-ray photographs were taken of both forms, oscillation photographs and Weissenberg photographs, but only the latter were used for the estimation of the intensities. For form *A*, only the $h0l$ reflexions were recorded on zero layer Weissenberg photographs, rotation axis b , with exposure times of 3, 6 and 12 hr. For form *B* all hkl reflexions were recorded. The main series were equi-inclination Weissenberg photographs, rotation axis b , zero and five layer line photographs. Two zero layer photographs were taken in addition to act as correlating photographs and showing $hk0$ and $3hkl$ reflexions respectively. Varying exposure times were used for all photographs showing strong reflexions.

Considerable difficulty was experienced in the photography of form *B* owing to the fairly rapid decomposition of the crystals which appeared to be hastened by exposure to X-rays. It was found necessary to use a different crystal for each individual Weissenberg layer line photograph in order to get comparable results. The sizes of the crystals were all of the order of $0.5\text{--}0.8 \times 0.2\text{--}0.4$ mm. The weights varied from approximately 0.05 to 0.10 mg.

The intensities of the X-ray reflexions were all estimated visually by comparison with a standard series of spots of varying exposure. Their relations were carefully checked owing to the large number of photographs taken, but it was realized that the irregular form of the crystals made it impossible to expect results of high accuracy. The intensities of the reflexions gradually faded out on the photographs taken, no reflexions with values of $\sin \theta/\lambda$ greater than 0.65 being observed. In all only 301 hkl reflexions were measured for form *B*.

To obtain a series of relative F^2 values the observed intensities were corrected by the method of Warren & Fankuchen (1941) for the Lorentz and polarization factors. An attempt was also made to apply a correction for absorption to the intensities of the $h0l$ reflexions in form *B*. The linear absorption coefficient μ of the crystal required in the expression $I = I_0 e^{-\mu d}$ was calculated from the formula $\mu = \frac{n}{V} \sum \mu_a$, where n is the number of molecules in, and V the volume of, the unit cell, and μ_a the atomic absorption coefficient. For cholesteryl iodide $\mu_{\text{calc.}} = 107$. The path distance, d , could only be found approximately through the procedure of projecting on to a sheet of graph paper the cross-section of the crystal normal to the axis of rotation and measuring path distances through the crystal from the centre of the cross-section. Here our chief difficulty was the irregular form of the crystal in cross-section. In the actual experiment a small crystal (weighing 0.05 mg.) was first

photographed on the Weissenberg camera to obtain the series of $h0l$ reflexions (rotation axis b), then removed and remounted to oscillate about the γ vibration direction. An oscillation photograph fixed the relation of the crystallographic axes to the rotation axis and consequently to the irregular crystal cross-section parallel to (010). A magnified image of this cross-section was then thrown on to a sheet of graph paper by means of a projector (see figure 1) and the path distances evaluated graphically.

The relative F values deduced for forms A and B are recorded in tables 2 and 3.*

(2) Calculation of Patterson projections

The first step in the structure analysis was the preparation of Patterson projections on the plane (010) for both cholesteryl iodide A and B using relative F^2 values for the $h0l$ reflexions. These Patterson projections are reproduced in figures 2*a* and 3*a*, and both show clearly recognizable peaks due to the iodine-iodine vectors. Since the y parameters of the iodine atoms in the structure may be assigned from crystal symmetry, the parameters of the iodine atoms were deduced as follows:

<i>Parameters of the iodine atoms</i>					
<i>A</i>			<i>B</i>		
$x = 0.043$	-0.043		$x = 0.217$	-0.217	
$y = 0.25$	0.75		$y = 0.25$	0.75	
$z = 0.248$	-0.248		$z = 0.042$	-0.042	

(3) Calculation of Fourier series using phase angles deduced from iodine contributions

(a) Fourier projections on (010) of electron density were then derived for both structures using the phase angles calculated from the contributions of the iodine atoms alone (figures 2*b*, 3*b*). In both these projections it is relatively easy to pick out the outlines of the sterol ring system and the side chain attached to it oriented as expected within the crystal. In the stable form A , the two molecules overlap throughout their length and this produces some confusion in projection, one ring being obscured by the iodine atom of the second molecule. In form B , on the other hand, the arrangement is more open and only a small amount of overlapping in projection occurs. The position of almost every atom in the sterol skeleton is fairly obvious here, and the structure determination was therefore continued for the time being on form B alone.

Since the projection (figure 3*b*) is centro-symmetrical, the phase angles calculated from the iodine contributions are 0 or π and are largely the correct phase angles. Very little refinement of the projection could be effected. The insertion of the calculated carbon atom contributions only changed the signs of three weak reflexions, 403, 508 and 509. The effect of the application of the absorption correction is shown in figure 3*c*. Some improvement but no very marked change in the contours resulted.

* Tables referred to by italic numbers are not printed but are available for reference at Oxford.

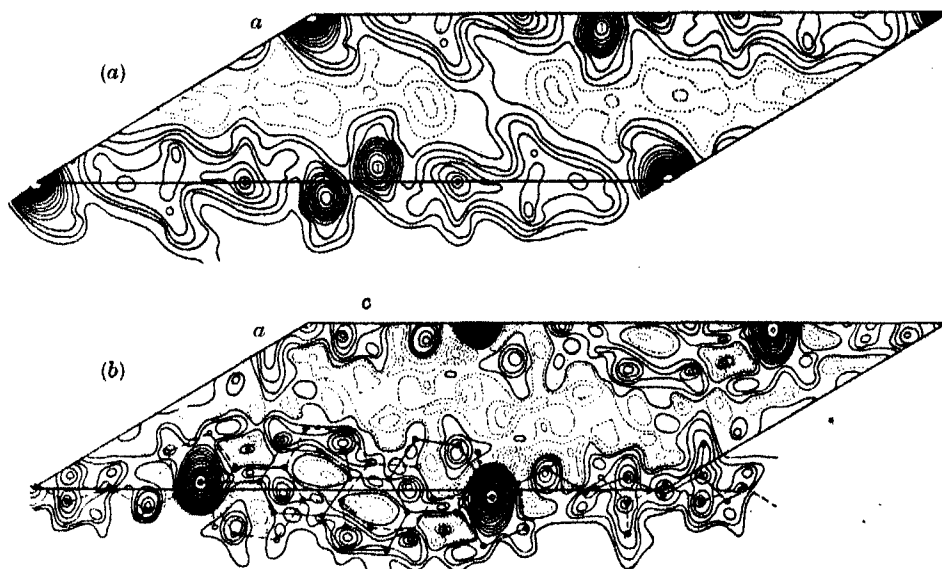


FIGURE 2. Cholesteryl iodide A. (a) Patterson projection (P_{xz}), on (010). The peaks I, I are due to the iodine-iodine vectors. (b) Projection of electron density, (ρ_{xz}), on (010). Lower contours dotted in these and succeeding figures to show contrast, and contour interval trebled over the iodine and origin peaks.

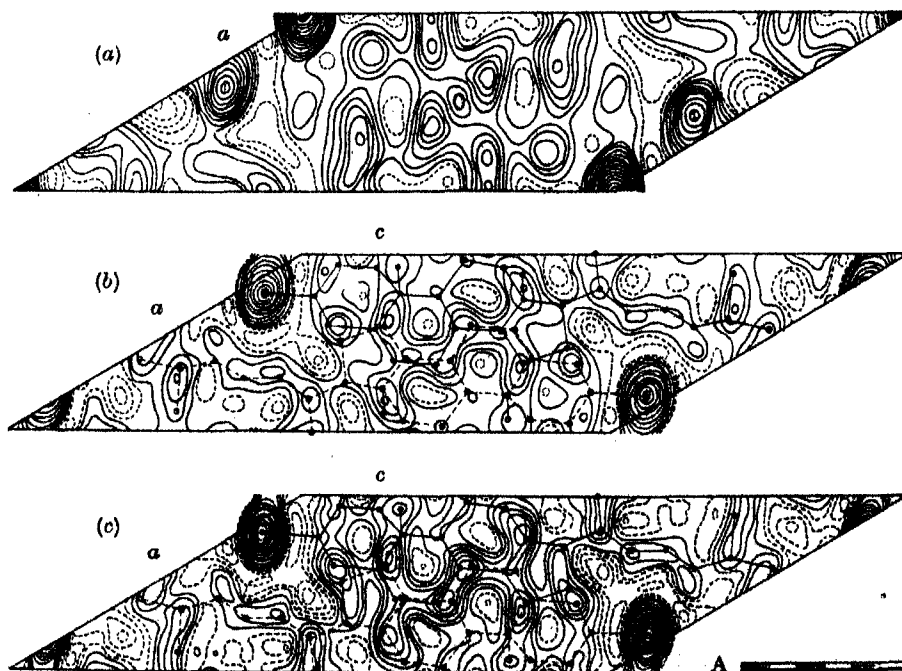


FIGURE 3. Cholesteryl iodide B. (a) Patterson projection (P_{xz}), on (010). The peaks I, I are due to the iodine-iodine vectors. (b) First projection of electron density, (ρ_{xz}), on (010). (c) Projection of electron density on (010) calculated, using intensities corrected for absorption and signs corrected for the carbon atom positions. Contours at rather closer intervals than (b).

(b) The next stage in the analysis is the calculation of three-dimensional Fourier syntheses along lines normal to the projection on (010), using all values of F_{hkl} and the phase angles calculated from the iodine contribution only. The x and z parameters for these lines were chosen from the projections so that they should if possible pass through atomic positions. But in attempting to fix these positions as closely as possible some 37 series in all were calculated (figure 4) which, as figure 5 shows, closely cover all the area of pronounced electron density in the projection. For convenience of reference these lines are numbered as far as possible according to the chemical formulation I and not in numerical sequence.

At this stage the phase angles from the iodine contributions are again 0 or π , and consequently the electron density pattern calculated by their use has a centre of symmetry. In addition to any particular real carbon atom P , which occurs at the positions (1) x, y, z ; (2) $\bar{x}, y + \frac{1}{2}, \bar{z}$ in the unit cell, the calculated pattern will show evidence of additional unreal atoms, (3) $\bar{x}, \bar{y}, \bar{z}$; (4) $x, \frac{1}{2} - y, z$, produced by the operation of this centre of symmetry. Of these, atoms 1 and 4 and atoms 2 and 3 are related to one another by mirror planes of symmetry at $y = \frac{1}{4}$ and $\frac{3}{4}$. We have found it most useful to consider the calculated pattern as showing the molecule as a whole superimposed upon its mirror image, produced by the operation of these mirror planes. The lines of electron density calculated (figure 4) are accordingly symmetrical about the planes $y = \frac{1}{4}, \frac{3}{4}$, and the y_p parameter of any particular carbon atom at x_p, y_p, z_p may be deduced as either $\frac{1}{4} + y'_p$ or $\frac{1}{4} - y'_p$.

At this point the direct X-ray analysis is at an end, and it is worth commenting on the bearing of the evidence so far obtained on sterol structure. The line syntheses of electron density have certain characteristics which should be noted:

- (i) Curves for carbon atoms 1, 2, 9, 10, 11, 12, 13, 16, 17, 18, 19, 20, 21, 22, 23, 25, and 26 show marked maxima mainly in the region $y = \frac{1}{4}$.
- (ii) Curves for carbon atoms 4, 5, 6, 7, 8, 14, 15, and 27 show maxima mainly in the region $y = \frac{1}{4}, y = \frac{3}{4}$.
- (iii) Curves for carbon atoms 18 and 19 show maxima in the region $y = 0.25, y = 0.42, y = 0.08$.
- (iv) The curves vary in form between two extremes. There are curves showing one more or less sharp maximum at $y = \frac{1}{4}$ (cf. 10) or two maxima at $y = \frac{1}{4} \pm y'$, where y' is generally less than 0.1 (cf. 1, 4, 12).

The first two characteristics serve to distinguish between the two molecules in the unit cell, since peaks separated by a distance $\frac{1}{2}b$ or 4.5 Å are too far apart to represent directly bonded atoms. The peaks at $y = 0.08, 0.42$ can be correlated with the presence of methyl groups attached at $C_{10}-C_{13}$ in the sterol skeleton. The fourth characteristic is an indication of the non-planar nature of the sterol skeleton. In general carbon atoms either lie not far from the plane $y = \frac{1}{4}$ or some little distance removed at $y = \frac{1}{4} \pm y'$, where the two possible positions have still to be distinguished. The parameters deduced are listed in table 4 and are necessarily approximate. This is because maxima may at this stage occur at some distance from the true position, owing to the incompleteness of the series and may also be confused through overlapping with the mirror image form.

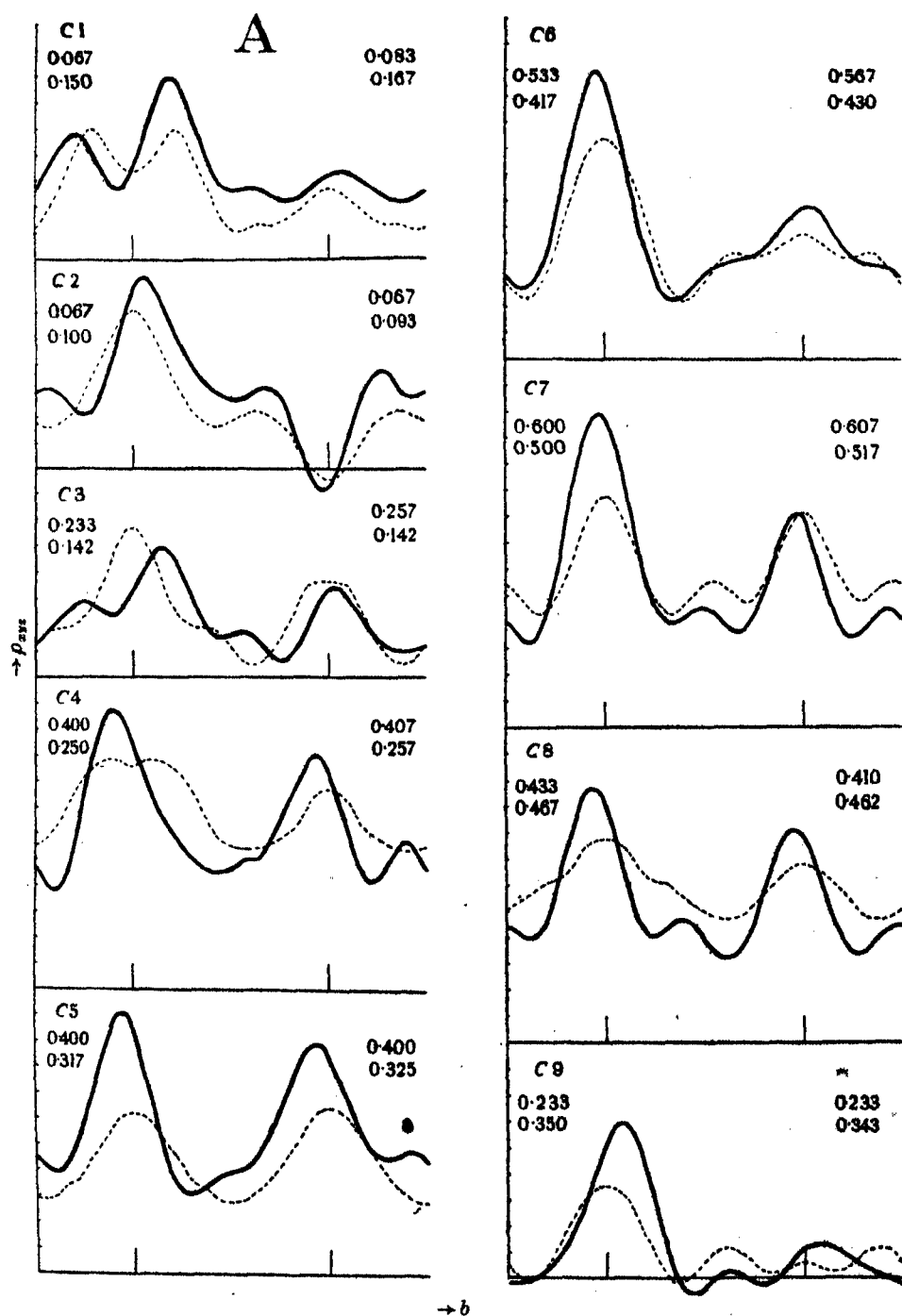


FIGURE 4. Cholesteryl iodide B. Curves showing electron density along lines parallel to b passing through atoms C_1 – C_9 . The dotted lines are the curves calculated using the phases due to the iodine atoms alone. For the full curves phases corrected for the carbon atom positions deduced were employed. The figures to the left refer to the x, z parameters for the dotted lines, those to the right for the full lines.

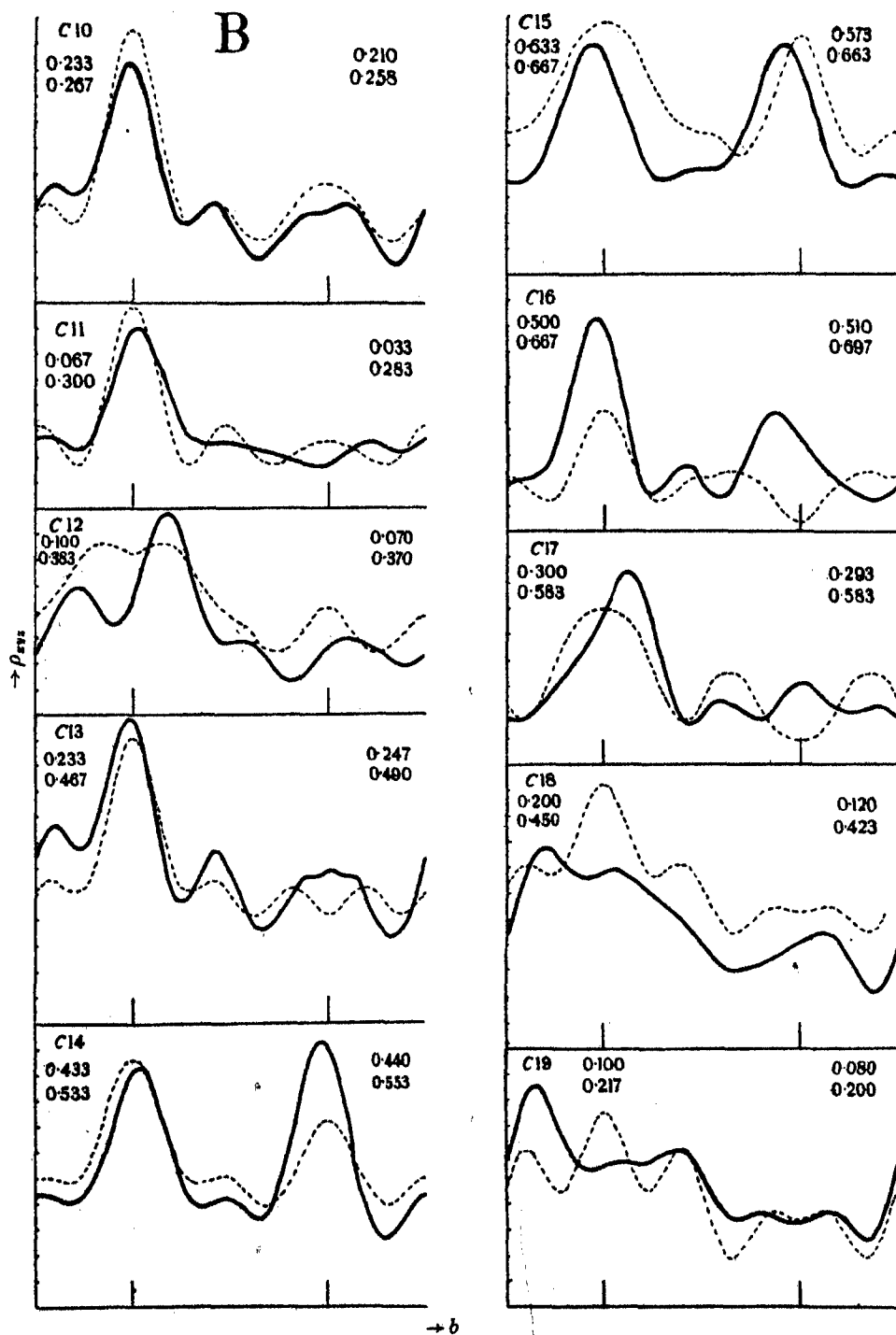


FIGURE 4 B

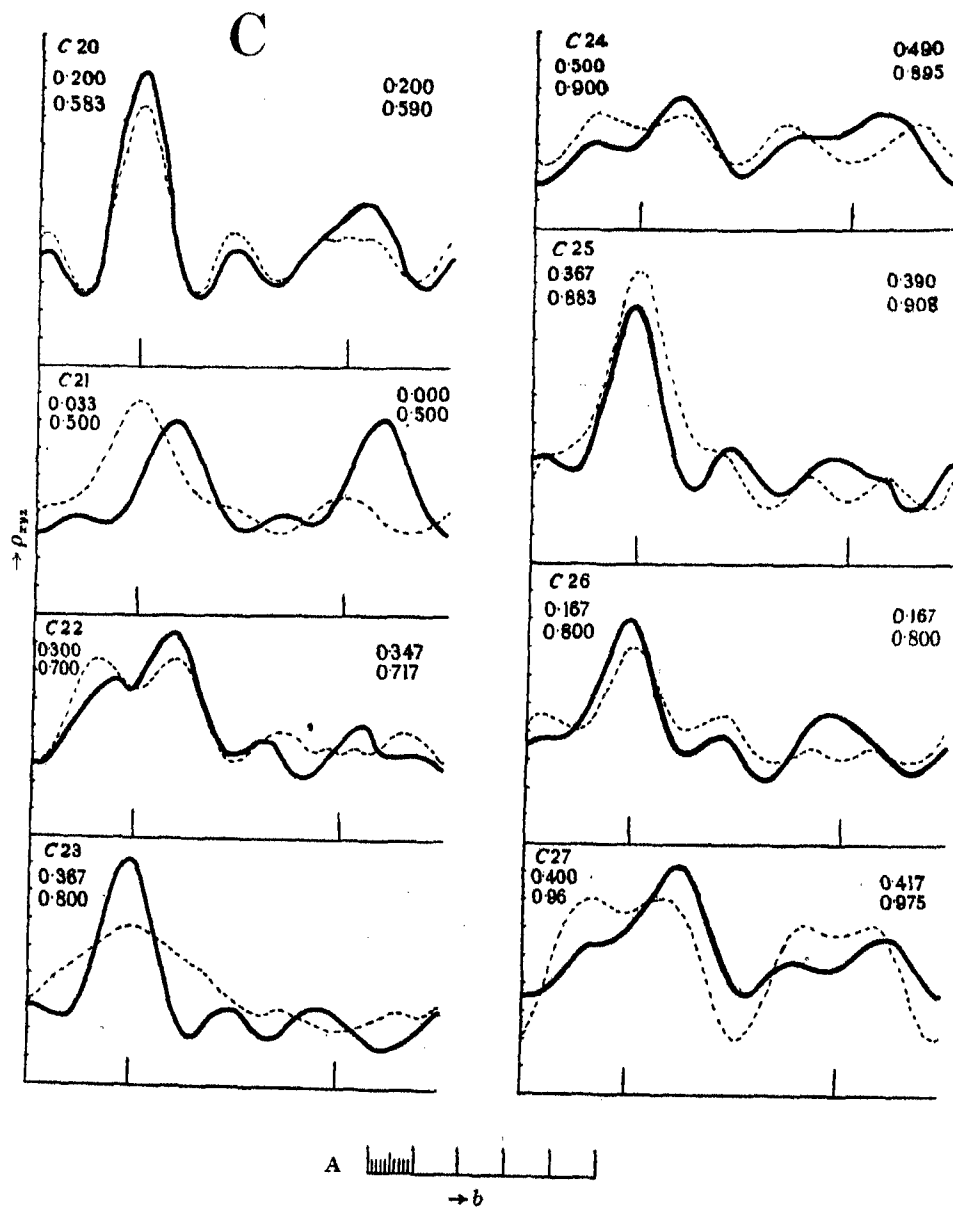


FIGURE 4 C

(4) *The choice of atomic positions*

In order to distinguish in three dimensions between the possible atomic positions deduced, use was made of our knowledge of normal carbon-carbon bond distances and valency angles. To assist us a scale model was set up as shown in figure 9 (plate 3). The unit cell, projected on (010), was plotted on a sheet of cork matting to a scale of 4 cm. = 1 Å, and at the atomic positions deduced from projection 3b, 28 vertical

steel needles were set up. On each of these needles were placed two pieces of cork above (+) and below (−) a plane corresponding to $y = \frac{1}{4}$ and at distances from this plane deduced from the corresponding line syntheses. The model then showed the approximate atomic positions of one molecule and of its mirror image. To separate

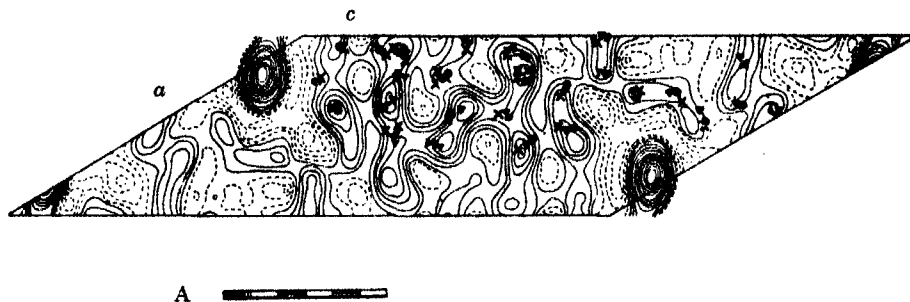


FIGURE 5. Cholesteryl iodide *B*. Electron density projection on (010) showing points through which line syntheses were calculated. \times Positions chosen for first (symmetrical) set of line syntheses. ∇ Positions chosen for second (asymmetrical) set of line syntheses. \bullet Atomic positions finally accepted.

the two a small cardboard triangle was constructed with two edges AB , BC (figure 6) proportional to 1.54 Å, and the included angle $109\frac{1}{2}^\circ$. The position of C_{10} at $y = 0.25$ was selected as first starting point and the position of C_1 arbitrarily fixed at $y = \frac{1}{4} + 0.09$ or $y = 0.34$. Then if one edge, AB , be placed along the line $C_{10}C_{11}$ the point C should fall on C_2 . This proved possible only if C_2 were placed at the position C_{2+} or at $y = \frac{1}{4} + 0.073$. From the adopted position of C_2 , the position of C_3 could be selected in the same way, and the process continued round the molecule. The chosen positions were checked by selecting other starting points and by moving round the individual rings in the opposite direction.

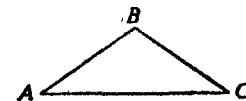


FIGURE 6. Triangular measure used in selecting the atomic positions.

It is perhaps worth illustrating the procedure by a direct calculation. Using the very approximate atomic parameters available at this stage the following bond distances and angles were deduced for the first atomic positions selected above:

$C_{10}-C_{1\pm}$	1.34		
$C_{1+}-C_{2+}$	1.44 Å	$\angle C_{10}-C_{1+}-C_{2+}$	114.6°
$C_{1+}-C_{2-}$	2.14 Å	$\angle C_{10}-C_{1+}-C_{2-}$	98.46

Position chosen C_{2+} . Actually the choice here is clearer than in the majority of cases (tables 6, 7).

Two modifications in the process described above were made. It was assumed in the first trial that, in ring *B*, a short interatomic distance $C-C = 1.34$ Å would be expected between C_5 and C_6 , and bond angles modified to equal 120° , and also that, in ring *D*, the bond angles might be considerably strained. In fact, as the calculations given illustrate, neither of these assumptions has much meaning since all bond

angles and distances deducible at this stage are very approximate. The atoms C_4 , C_5 , C_6 , C_7 all appear to lie so near the plane $y = \frac{1}{4}$ that no great error should be introduced by misplacing them. More difficulty was experienced in the neighbourhood of the ring junction $C:D$, though check calculations do favour the configuration first chosen from the model (tables 6, 7).

The atomic positions selected as a result of this sorting process on the model are shown in the photograph (figure 9, plate 3), by the shaded corks. Since these represent very approximate versions of the actual atomic parameters in the crystal a further model was constructed, having the stereochemical form actually deduced but theoretical bond distances and valency angles. This was used to amend slightly the first set of atomic parameters chosen within the limits of error provided by the Fourier series so far calculated. The parameters adopted are listed in table 5*a*.

(5) *Refinement of the structure through calculations of three-dimensional Fourier series using corrected phase angles*

At this stage the approximate position of every atom in the molecule was known and the phase angles were therefore calculated, using both the carbon atom and iodine atom contributions. With the observed F_{hkl} values and these phase angles new three-dimensional Fourier series were derived of two main types: line syntheses parallel to $[010]$ and sections normal to $[010]$.

The three-dimensional line syntheses, figure 4 (continuous lines), show electron density along lines parallel to $[010]$. They are, as expected, no longer symmetrical about $y = \frac{1}{4}$, $\frac{3}{4}$ and consequently permit far more accurate determination of the y parameters. As expected, too, the main peaks have in certain cases shifted quite markedly from the position first selected but in no case in a direction to change the molecular configuration chosen. Usually small peaks also occur at the mirror image positions indicating that too much weight has still been given to the iodine contributions.

The sections show electron density in planes parallel to the plane (010) , figure 5. Their heights above the plane were chosen to pass through as many atomic positions as possible in order to obtain better values for the x and z parameters. Section (a) at $y = \frac{1}{4}$ passes through, to a greater or less degree, 17 out of the 27 carbon atoms of the sterol skeleton. It shows particularly well the five-membered atoms of ring D . Section (b) at $y = 0.33$ is complementary to this. It does not in general pass so near the atomic positions (cf. table 5). Parts only of section (c) at $y = 0.12$ were calculated to cover the two remaining atoms, the methyl groups at C_{18} and C_{19} .

The electron density peaks representing atoms in these three-dimensional series are in many cases not sharp and, as the line series illustrate particularly well, often not equivalent in strength as they should be theoretically. The series also show considerable spurious detail which can be largely correlated with the atomic positions of the mirror image molecules. It is clear that the atomic parameters and consequently phase angles used for their calculation were only approximately correct and are still too close to those given by the iodine atoms only. The same point is

illustrated by the calculation of F values listed in table 3. There is an apparent general agreement which depends largely on the values of the iodine contributions, but a more detailed examination of the carbon atom contribution shows several discrepancies. From the three-dimensional series, however, a number of improvements may be made in the atomic parameters, the revised values of which are listed

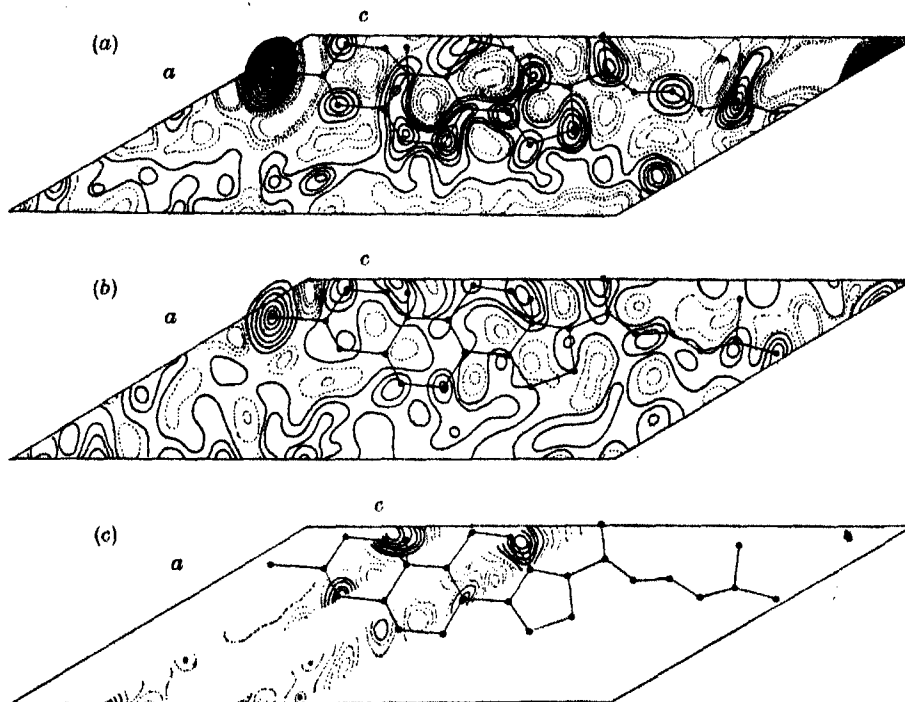


FIGURE 7. Cholesteryl iodide *B*. Sections showing electron density in planes parallel to (010) at heights $y = 0.25$ (a), 0.33 (b), 0.12 (c).

in table 5*b*. We should proceed next to recalculate phase angles and then electron density employing these parameters. But the calculations involved in the derivation particularly of the sections are so exceedingly lengthy that we have felt it best to give the results as they stand now without attempting at present further refinement of the structure. There is an additional reason for this in that from this point any improvement in the structure determination rests mainly on trial and error analysis.

DISCUSSION

(a) Crystal structures of cholesteryl iodide *A* and *B*

The crystallographic view of the structures of the two forms of cholesteryl iodide founded on the preliminary measurements has been remarkably closely confirmed by this more detailed investigation. In both the molecules appear as essentially long

with the side chain extending from and continuing the line of the ring system. They are also roughly flat, the plane of the ring system being, as predicted, nearly parallel to the b plane. Once the actual atomic positions in a molecule of this complexity have been deduced, rough calculation of size and shape or inclination of the ring system to the crystal axes largely lose their meaning. For interest, however, the following figures may be given:

	thickness	width	length
dimension of sterol skeleton first given by J. D. Bernal (1932)	5.0	7.2	17-20
'dimensions' of cholesteryl iodide	4.3	7.5	21.4

(Here thickness, height of C_9 above C_8 , width, C_7-C_{11} , length $C_{17}-I$, all $+3.5$ Å. Inclination of ring system: line $C_6-C_7-10^\circ$ to b plane.)

The orientation of the crystallographic optic axes to the molecular axes is also of interest in this connexion. The direction of least refractive index, α , is parallel to b and, as expected, nearly normal to the plane of the ring system. In this particular orientation of the sterol molecule to the crystal axes the direction of largest refractive index, γ , could not in the preliminary examination be correlated with any certainty with the length of the molecule. In fact, owing to the overlapping molecular arrangement in form A , γ does here lie very closely along the direction of greatest molecular length. In form B there is no such marked concentration of atoms along any one line in the plane (010) and correspondingly the positive character of the crystals appears less pronounced (compare the optic axial angles measured—for form A , $2V = 56^\circ$, for form B , $2E = 88^\circ$). The direction of greatest atomic density is no longer that of the molecular length; γ accordingly diverges from this (figure 8). It is noticeable that γ is here also nearly parallel to the direction of the double bond and, without calculation, it is impossible to say whether this or the electron density principally determines its orientation.

The relation between the molecular arrangements in the two forms is very close (figure 8). Within a single layer parallel to the b plane the relative arrangements of the molecules are nearly, though not quite, identical. The two crystal structures differ principally in the relative arrangement of the layers. In B there is an open arrangement; the only overlapping of the atoms in the layers occurs where it might be expected, on the side of the molecule free from projecting methyl groups. The crystal structure of A is derived from B by sliding one layer over the next through a distance of roughly 4.7 Å into more closely packed positions. A small readjustment of the atomic position within the layers follows—the plane of the ring system is probably at a slightly greater inclination to the b plane in A than in B and a rotation of C_{20} about the bond $C_{17}-C_{20}$ appears to have taken place, bringing C_{21} almost vertically below C_{20} in the b plane projection. There are one or two carbon-carbon distances of about 3.5 Å but the greater number, particularly in form B , are 4 Å or more. On the other hand, in form B , the two iodine atoms are much closer together than they are in form A , in fact only 6.1 Å apart as against 11.9 Å in A . It

seems reasonable to correlate this feature of the crystal structure with the deeper colour and more ready decomposition of the crystals of *B* which distinguish them at first sight from those of *A*.

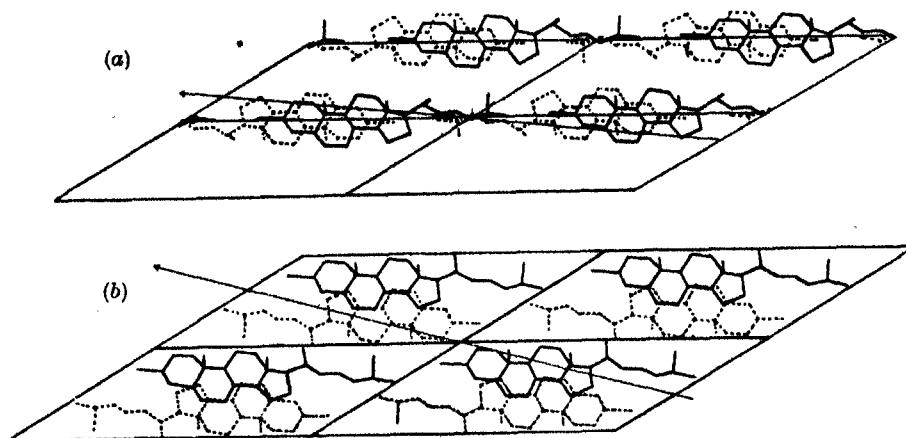


FIGURE 8. Diagram to illustrate the relation between the crystal structures of cholesteryl iodide *A* (a) and *B* (b). The molecular positions are shown projected on (010).

(b) *The accuracy of the detailed structure determination*

The principal inaccuracies in the detailed structure determination of cholesteryl iodide are certainly a consequence of the incompleteness of the calculations combined probably with some inaccuracies in the intensity measurements. In addition there are certain other features of the data we have which make it difficult to fix atomic positions exactly.

The first of these is the occurrence of specific diffraction effects as a result of the presence of the iodine atom (cf. Robertson & Woodward 1940). This probably accounts for the comparative weakness of the peak due to C_3 , particularly as this weakness appears also in the projection on (010) which is least likely to be affected by errors in the phase angles and intensities. As a result some doubt was felt about the exact placing of C_3 .

The second feature is the general weakness of X-ray reflexions from planes with small spacings, which seriously limits the series used and consequently the resolution of the electron density peak system. This effect indicates some disorder in the crystal structure, i.e. that the atomic positions are not in fact precisely fixed in space. It seems most reasonable to correlate this with the complexity of the molecular system present. Where, as here, complicated molecules are held in the crystal by comparatively weak intermolecular forces it is unlikely that the atoms will be so precisely ordered throughout space as they are in simpler crystal structures. An interesting commentary on the situation is provided by the comparison of the crystal structures of cholesteryl iodide *A* and *B*, which show varying but closely related stable arrangements of the molecules.

As a result of these effects accurate measurements of interatomic distances and valency angles cannot be expected from our present data. But the X-ray analysis has very largely revealed the structure of the molecule in the sense used in most researches in organic chemistry. It is possible to distinguish the different molecules in the crystal and to determine which atom is attached to which within the molecule and in addition to fix with a considerable degree of confidence the relative orientations of the atoms in space. The evidence on molecular structure is the more reliable here in that the detailed information obtained from cholesteryl iodide *B* is confirmed by the data we have on cholesteryl iodide *A*.

(c) *The molecular structure of cholesteryl iodide*

The molecular structure of cholesteryl iodide derived in the present analysis is defined by the set of parameters found for all the atoms in the crystal of cholesteryl iodide *B* and listed in table 5*b*. The meaning of these figures in chemical terms is shown more clearly by the photograph of a model (figure 10, plate 3) built up on the crystallographically deduced atomic positions. In general form this molecular

TABLE 5*b*. THIRD SET OF APPROXIMATE PARAMETERS FOR
CHOLESTERYL IODIDE, FORM *B*

atom	<i>x</i>	<i>y</i>	<i>z</i>	atom	<i>x</i>	<i>y</i>	<i>z</i>
I	0.217	0.250	0.042	C ₁₄	0.443	0.260	0.550
C ₁	0.083	0.347	0.167	C ₁₅	0.685	0.220	0.681
C ₂	0.045	0.280	0.083	C ₁₆	0.513	0.230	0.694
C ₃	0.240	0.323	0.142	C ₁₇	0.308	0.307	0.585
C ₄	0.400	0.200	0.249	C ₁₈	0.117	0.100	0.422
C ₅	0.423	0.220	0.328	C ₁₉	0.123	0.083	0.222
C ₆	0.597	0.227	0.443	C ₂₀	0.193	0.250	0.585
C ₇	0.607	0.233	0.520	C ₂₁	0.000	0.333	0.492
C ₈	0.417	0.220	0.469	C ₂₂	0.313	0.333	0.695
C ₉	0.233	0.293	0.342	C ₂₃	0.307	0.240	0.752
C ₁₀	0.223	0.243	0.272	C ₂₄	0.407	0.340	0.850
C ₁₁	0.053	0.260	0.297	C ₂₅	0.365	0.237	0.884
C ₁₂	0.070	0.333	0.370	C ₂₆	0.137	0.237	0.780
C ₁₃	0.237	0.240	0.490	C ₂₇	0.450	0.367	0.968

structure is in good agreement with present chemical views (Rosenheim & King 1932; Wieland & Dane 1932). The ring system of three six-membered rings *A*, *B* and *C* and one five-membered ring *D*, is non planar. The structure of the side chain is clear even from the first Fourier projection, and there seems no doubt that the methyl groups are attached at C₁₀ and C₁₃. The interatomic distances and valency angles calculated for this model also all agree reasonably well with theory. They are given in tables 6 and 7 and may be summarized here:

C—I	2.08		
C—C	1.47–1.80	Mean	1.55
C=C	1.30		
∠C—C—C	91°–129° 30'	Mean	108° 36'
∠C=C—C	124° 45', 125° 33'		

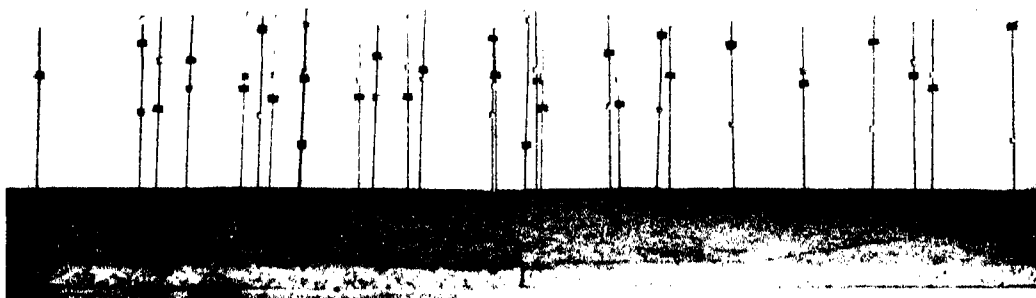


FIGURE 9. Trial model of cholesteryl iodide showing the positions of the atoms of the molecule and its mirror image (black and white corks).

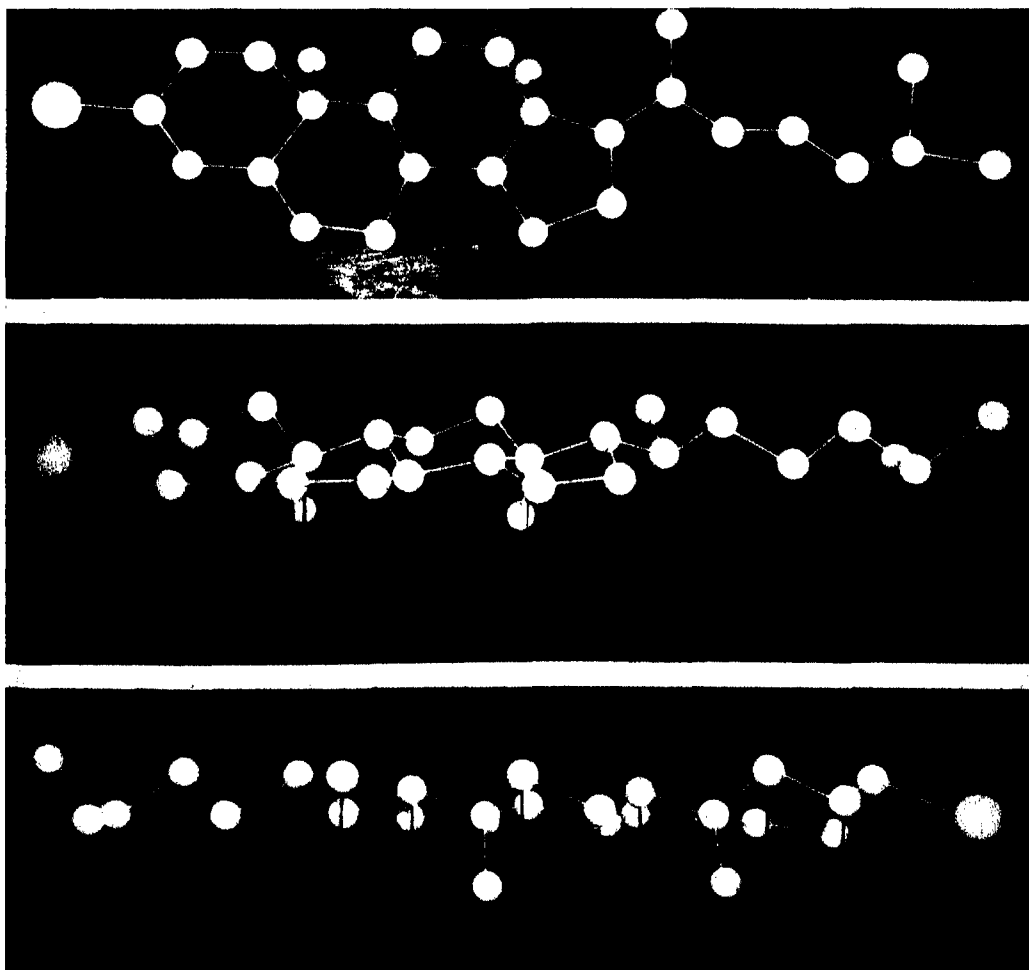


FIGURE 10. Photographs of the final model of cholesteryl iodide.

(Facing p. 80)

Since the accuracy is not great, it is unlikely that any of the deviations from normal bond distances are significant. The same is true for most of the valency angles found, but the most marked deviations do occur near the double bond and within and adjacent to the five-membered ring; these are probably largely real since in these regions some strain must be present.

The stereochemical configuration of rings *A* and *B* of cholesteryl iodide are largely as would be expected from chemical theory. It is unfortunate that the position of C_3 is so ill-defined in the electron density maps, but the data so far obtained definitely favour the position given.* The configuration of the carbon-iodine bond is here *cis* to methyl at C_{10} (the so-called '*trans*' form of Ruzicka, Furter & Goldberg 1938). This configuration is the same as that most favoured on chemical grounds for the hydroxyl group in cholesterol, and is therefore in general agreement with other evidence that no inversion occurs in the formation of the cholesteryl halides from cholesterol (Bergmann 1937, 1938). The most interesting feature of the *A*, *B* ring system is the distortion that results from the presence of the double bond at C_5-C_6 . The atoms C_4 , C_5 , C_6 , and C_{10} all lie very nearly in one plane. The effect is shown even in the first three-dimensional line series. These accordingly provide some independent evidence for placing the double bond among the atoms concerned; though it may be doubted whether the actual double-bond position could have been deduced with any certainty directly from the X-ray data. Apart from the distortion caused by the double bond the rings have the Sachse *trans* configuration.

As would be expected ring *C* of the sterol skeleton most nearly approaches the regular Sachse *trans* form. The configuration of the carbon atoms about the bonds C_8-C_9 , C_8-C_{14} and $C_{13}-C_{14}$ are all *trans*. The first involves the junction between rings *B* and *C*, and here the present evidence is good (cf. Ruzicka & Thomann 1933). The last concerns the junction between rings *C* and *D* (Wieland & Dane 1933; Dimroth & Jonsson 1941). Here there must in any case be some strain in the system, and it was the less easy at stage 4 of the analysis to be certain that the correct atomic position had been chosen. All that can be stated is that the present data favour the *trans* configuration. This involves a small but definite distortion of the five-membered ring. The four atoms C_{13} , C_{14} , C_{15} , C_{16} lie nearly in one plane with C_{17} a short distance from this plane. All the calculated valency angles in the ring are smaller than 108° .

So far the detailed X-ray analysis has largely confirmed points on which there is also chemical evidence. On the stereochemical relations involved in the attachment of the side chain at C_{17} the chemical evidence is either conflicting or non-existent (Ruzicka, Goldberg & Wirz 1935; Wieland & Dane 1933*b*). Here the crystallographic evidence is the more reliable in that it rests principally on the projection

* Change to the alternative chemical configuration requires change in the y parameters of both C_3 and C_4 . In the first three-dimensional series C_3 was very near $y = \frac{1}{2}$, and it has moved nearer the position expected from its apparent relation to C_3 and C_4 in the second series. This is some confirmation that the correct relation of the three has been selected.

on (010). It is strongly in favour* of the attachment of the side chain *cis* to methyl at C₁₃ as suggested in the earlier work of Caglioti & Giacomello (1939). The arrangement of the carbon atoms about the bond C₁₃-C₁₇ is then again *trans* in form and the same is true of the bond C₁₇-C₂₀. It is an interesting point that in this region, and indeed throughout chain and ring systems of the molecule, the arrangement of the carbon atoms largely conforms with the staggered *trans* configuration, the stability of which has been demonstrated in other complex organic structures particularly by C. W. Bunn (1942). This may indeed be the characteristic of the structure which favours the otherwise apparently less stable configurations of the junction between rings *C* and *D*.

Our thanks are due to the Rockefeller Foundation for a grant towards X-ray apparatus and to the Department of Scientific and Industrial Research and the Trustees of the Carnegie Research Fund for research grants to one of us (C.H.C.). Much of the apparatus used was evacuated to Oxford from Birkbeck College, London.

REFERENCES

- Bergmann, E. 1937 *Helv. Chim. Acta*, **20**, 590.
 Bergmann, E. 1938 *J. Amer. Chem. Soc.* **60**, 1997.
 Bernal, J. D. 1932 *Nature*, **129**, 277.
 Bernal, J. D. & Crowfoot, D. 1933 *Trans. Faraday Soc.* **29**, 1032.
 Bernal, J. D., Crowfoot, D. & Fankuchen, I. 1940 *Phil. Trans. A*, **239**, 135.
 Beynon, J. H., Heilbron, I. M. & Spring, F. S. 1936 *J. Chem. Soc. p.* 907.
 Bragg, W. H. 1915 *Phil. Trans. A*, **215**, 253.
 Bragg, W. L. 1929 *Z. Kristallogr. A*, **70**, 475.
 Bunn, C. W. 1942 *Proc. Roy. Soc. A*, **180**, 67.
 Caglioti, V. & Giacomello, G. 1939 *Gazz. chim. ital.* **69**, 245.
 Carlisle, C. H. 1943 A crystallographic investigation of the stereo-chemistry of the sterols and related compounds. Thesis for D.Phil. degree. Oxford.
 Cox, E. G. & Jeffrey, G. A. 1939 *Nature*, **143**, 894.
 Dimroth, K. & Jonsson, H. 1941 *Ber. dtsch. Chem. Ges.* **74**, 520.
 Goodwin, T. H. & Hardy, R. 1938 *Phil. Mag.* **25** (vii), 1096.
 Harker, D. 1936 *J. Chem. Phys.* **4**, 381.
 Helferich, B. & Gunther, E. 1939 *Ber. dtsch. Chem. Ges.* **72**, 338.
 Huse, G. & Powell, H. M. 1940 *J. Chem. Soc. p.* 1398.
 Koechlin, B. & Reichstein, T. 1942 *Helv. Chim. Acta*, **25**, 918; cf. Reichstein, T. & Shoppee, C. W., *Vitamins and Hormones*, **1**, 350.
 Patterson, A. L. 1935 *Z. Kristallogr. A*, **90**, 517.

* Koechlin & Reichstein (1942) have quoted the fact that bisnordesoxy cholic acid does not lactonize, as possible evidence against this configuration providing the following assumptions are made: (a) OH at C₁₃ is *cis* to Me at C₁₃ and (b) lactonization will only be inhibited if the side chain is attached at the opposite side of the ring system to OH at C₁₃. Models show that —OH at C₁₃ *cis* to Me at C₁₃ is within easy reach of a —COOH at C₂₀ with either configuration at C₁₇-C₂₀. Inhibition of lactonization appears more likely with the *cis* configuration as a result of the relative positions of Me at C₁₃ and at C₂₄; it is also possible that the configuration of the OH group should be reversed since this rests on the evidence of Caglioti & Giacomello, and it is very difficult to be certain of the position of a single atom in a trial and error Fourier analysis of the complexity of that carried out by these authors.

- Robertson, J. M. & Woodward, I. 1940 *J. Chem. Soc.* p. 36.
Rosenheim, O. & King, H. 1932 *Chem. and Ind.* **51**, 464.
Ruzicka, L., Furter, M. & Goldberg, M. W. 1938 *Helv. Chim. Acta*, **21**, 498.
Ruzicka, L., Goldberg, M. W. & Wirz, H. 1935 *Helv. Chim. Acta*, **18**, 998.
Ruzicka, L. & Thomann, G. 1933 *Helv. Chim. Acta*, **16**, 221.
Warren, B. E. & Fankuchen, I. 1941 *Rev. Sci. Instrum.* **12**, 90.
Wieland, H. & Dane, E. 1932 *Hoppe-Seyl. Z.* **210**, 268.
Wieland, H. & Dane, E. 1933 *Hoppe-Seyl. Z.* (a), **216**, 91, (b) **216**, 98.
-

The electrical conductivity of stannous sulphide

BY J. S. ANDERSON AND (MRS) M. C. MORTON

Department of Chemistry, University of Melbourne

(Communicated by Sir David Rivett, F.R.S.—Received 26 June 1944)

The conductivity of stannous sulphide has been measured between the ordinary temperature and 400° C, and follows the expression $\kappa_T = A_1 e^{-E_1/kT} + A_2 e^{-E_2/kT}$. The low-temperature conductivity has been studied with particular reference to the effect of chemical treatment calculated to change any deviations from ideal stoichiometric composition. Treatment with hydrogen at temperatures below those at which reduction to tin can occur brings about a considerable decrease in conductivity; exposure to oxygen or to hydrogen sulphide restores the conductivity. The low-temperature conductivity is attributable to a stoichiometric excess of non-metal in the lattice, the current being transported by a corresponding number of positive holes. From the rate of diminution or restoration of conductivity, the positive holes must be located in the surface layers of atoms of the crystallites. The changes in conductivity resulting from changes in the stoichiometric defect arise almost entirely through the change in the activation energy of the conduction process, E_1 . Although not considered in the present theory of semiconductors, both E_1 and the electronic mobility are functions of the concentration of conducting centres.

The high-temperature conductivity is attributed to the intrinsic conduction of the lattice, and indicates the existence of a conduction band about 1.2 eV above the full band.

Most of the considerable body of observations, amassed in recent years, as to the properties of semiconducting solids has related to the metallic oxides, i.e. to compounds of predominantly ionic character. By contrast, few investigations have been made of the sulphides, selenides, etc., in which the magnitude of the polarization effects leads, as is generally conceded (cf. Mark 1940), to a marked transition of the lattice forces towards homopolar binding. This transition is evidenced by the characteristic properties (colour, metallic reflex, insolubility in ionic or ionizing liquids, etc.) of the chalcogenides of the heavy metals. Such compounds also frequently show a marked variability in composition to which, rather than intrinsic conductivity, their strikingly high electrical conductivity at room temperature is due.

We describe here experiments carried out on stannous sulphide. This compound, of strongly marked semi-metallic character, is (like PbS, studied by Hintenberger (1942) derived from the lower of two potential valency states of the metal, and two

possibilities suggest themselves in consequence: (a) that the unutilized valence electrons might lie close enough to the conduction band of the lattice to confer a marked intrinsic conductivity at low temperatures (cf. the last section of this paper), (b) that the potential quadrivalency of the metal should favour the formation of non-stoichiometric stannous sulphide, of subtractive solid solution type, which would be a deficit, or positive hole conductor. Moreover, stannous sulphide (m.p. 870°C) is one of the few sulphides which can conveniently be investigated in the fused state. An extension of the work in this direction is very desirable, because of our lack of knowledge of such liquids. Apart from their general interest, the constitution and properties of fused sulphides are of prime importance for a number of problems of theoretical geochemistry and metallurgy.

As will be seen from the experimental results, stannous sulphide does behave according to the rule of Friederich (1925) and Meyer (1933), which enunciates that binary compounds of the lowest valency state of a metal show enhanced conductivity by incorporation of excess non-metal while those of the highest valency state show enhanced conductivity by elimination of non-metal. Stannous sulphide displays a range of composition which, although too small to be detected analytically, is particularly marked on the sulphur-rich side of the ideal formula, and it is normally a deficit conductor. Its specific conductivity at low temperatures is a function of the composition, and can be varied within wide limits, whilst at higher temperatures the intrinsic conductivity of the lattice prevails.

After the main bulk of our work was completed, that of Hintenberger (1942) on lead sulphide became known to us. There is a close parallelism between the behaviour of the two compounds.

THE NATURE OF THE CONDUCTING PROCESS

Published information about the electrical properties of stannous sulphide is scant. Hittorf (1851) concluded that the conductivity was electrolytic, as also did Guinchant (1902) and Trey (1925). The latter based his view on the isolated observation that the resistance of a polycrystalline pellet was greater in one direction than another, but possible asymmetry in the contact resistances, and the occurrence of rectification, were ignored. Tubandt & Haedicke (1927), in one electrolytic transport experiment carried out by Tubandt's carefully developed technique, found no detectable electrolytic component of the conduction at 260°C .

This last result appears trustworthy. It seemed desirable, however, to extend Tubandt's measurements to higher temperatures. Electrolytic conduction in solids involves the formation and migration of lattice defects of Schottky or Frenkel types, and as the potential barriers between adjacent lattice sites or interlattice positions are usually high, the ionic conduction process in most solids is associated with a high activation energy. Qualitative observations on the speed of reaction of metallic tin with sulphur suggest that the mobility of the tin atoms in stannous sulphide is low; nevertheless, at elevated temperatures some 'platzwechsel' processes must

occur, and a fraction of the electric current may be transported by the ionic mechanism.

To set an upper limit to this electrolytic component of the conduction, we carried out a number of transport experiments by Tubandt's technique, at temperatures up to 750° C, in which the current was passed through a column of compressed pellets of stannous sulphide, clamped between electrodes of spectrographic graphite. As, owing to a number of experimental difficulties (especially the volatility of stannous sulphide), no very precise results were obtained, details of the experiments are not reproduced. It was, however, established that at 750° C the electrolytic conduction was certainly less than 1 % of the whole, and was probably immeasurable. Accordingly, stannous sulphide may be taken to be an electronic conductor at all temperatures. Direct confirmation that the current carriers at ordinary temperatures are positively charged is referred to in a later section.

EXPERIMENTAL

Our main series of experiments consisted of conductivity measurements, over the temperature range 15–500° C, carried out on pressed powder pellets subjected to varied chemical and thermal treatment. From these, the relation of the activation energy of the conduction process, the number of conducting centres and the sign of the current carriers to the stoichiometric composition and past history of the material may be examined.

Preparation of material

In view of the role played by impurities in the properties of semi-conductors, stannous sulphide was made by direct union of the purest available tin and sulphur. The tin was a sample of practically spectroscopic standard quality, kindly presented to us by the smelters, Messrs O. T. Lempriere Ltd., of Melbourne. This was heated (cf. Anderson & Ridge 1943) with Schuchardt's 'pure recrystallized' sulphur that had been subjected to a further process of vacuum fractionation. The product was homogenized by repeated crushing and reheating, and the crystalline product was finally freed from excess of sulphur by prolonged heating in hydrogen at 400–500° C until no further formation of hydrogen sulphide was detectable. Material from several different preparations showed general agreement in properties. Some samples were subjected to further treatment by fusion or sublimation, as indicated in the tables of data.

Stannous sulphide pellets. Polycrystalline pellets were prepared by compression in a cylindrical die, 0.82 cm. in diameter, under a pressure, applied by a screw press, estimated at about 6 tons per sq.in. The die and plungers were of hardened steel, and the ends of the plungers were precision-ground plane and square to the axis after hardening.

The pellets so obtained were lustrous and reasonably hard. Although too friable to be turned or polished in the lathe, they were strong enough to be handled without

any loss in weight, and could be strongly compressed without crushing. From their density and that recorded elsewhere for stannous sulphide, they consistently contained 10–20 % of intergranular void space. On heating in hydrogen or pure nitrogen at 550° C, recrystallization and sintering occurred, the pellets becoming notably harder and stronger without change in external dimensions or space-filling. The pellets used were 0.5–1.9 cm. in length, and had resistances of 40–400 ohms at 18° C in their initial state.

The use of polycrystalline pellets for conductivity measurements has been criticized by a number of workers (e.g. Streintz 1902; Koenigsberger 1910, 1914; Frey 1930; Fritsch 1935) on the grounds that unknown high intergranular contact resistances may yield quite false values for the conductivity. As against this, in one instance where comparable results are available for compressed pellets and compact macrocrystalline material, Fischbeck & Dorner (1929) report very fair agreement between different specimens of CuS. There is no alternative to the use of polycrystalline pellets if the composition and thermodynamic lattice equilibrium of the material are to be closely controlled. The large crystals of many compounds that occur as native minerals are invariably too impure to provide significant results, while large crystals of pure oxides, sulphides, etc., can be made only by fusion or sintering at high temperatures. It is noticeable from the literature that, in all work on fused and sintered specimens (e.g. the work of Fritsch on ZnO, and of Frey on PbS, cited above), abnormally high conductivities have been found, coupled with low values of the temperature coefficient and activation energy of conduction. From observations of our own on the fusion of stannous sulphide, and from the work of Frey, it would seem that two potential sources of abnormally high conductivity are introduced. First, with compounds of non-stoichiometric character, fusion may bring about a change of chemical composition. Thus, we have observed that pure stannous sulphide, after fusion in a vacuum, leaves a minute residue of metallic tin when it is sublimed, due to decomposition or disproportionation: below the melting-point such changes are limited by the slow rate of diffusion through the lattice. Secondly, at the melting-point a relatively high concentration of lattice defects will be present in thermodynamic equilibrium, and many of these may be 'frozen in' during the solidification and cooling processes. Data quoted by Meyer & Neldel (1937) for a sample of TiO_2 , in which a very high degree of lattice disorder was produced, may be cited in illustration of these considerations.

Apparatus. The pellets of stannous sulphide were mounted between electrodes of spectrographic graphite, the faces of which had been turned truly square and smooth in the lathe. The whole assembly was then clamped in a small metal press (A, figure 1). To eliminate the contact resistance between the carbon plate and the SnS, the press was progressively tightened until the over-all resistance decreased to a constant value, which remained unchanged on further increase of pressure, and which was reproducible if the whole were dismantled and reassembled. Having regard to this fact, to the consistent values for the conductivity obtained with pellets varying in length (and therefore in total resistance) by a factor of 4 to 1, to

the strictly linear nature of the $\log R - 1/T$ curves, and to the lowest values of resistance measured in our experiments (about 1 ohm, at higher temperatures), we consider that significant contact resistances did not remain. Miller (1941) reported contact resistances with the same temperature coefficient as the bulk resistance of his zinc oxide specimen; if such were present in our work they would serve only to shift the measured conductivity curves bodily by a small amount. The thermal expansion of the SnS pellets appeared to be somewhat greater than that of the metal press, so that no diminution of the clamping pressure occurred at higher temperatures.

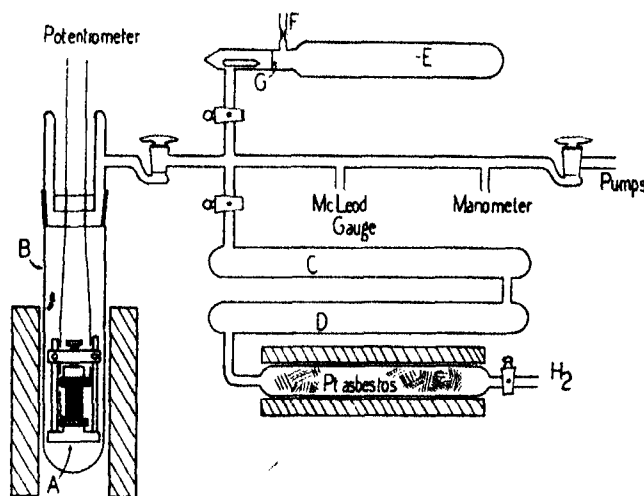


FIGURE 1

The press was contained within the tube *B*. Electrical connexions to the electrodes were made by means of nickel wires, and were carried out of the apparatus through a tungsten-pyrex pinch on the upper member of the ground joint. Tube *B* was connected to a vacuum line, with provision for admitting hydrogen or pure hydrogen sulphide, both rigorously free from traces of oxygen. The temperature of the specimen was controlled by an electric furnace surrounding the tube *B*, and was measured by a calibrated chromel-alumel couple.

Hydrogen. Commercial electrolytic hydrogen, containing 0.2 % of oxygen, was used. This trace of oxygen was removed by passing the gas through heated platinized asbestos, and the water so formed was absorbed in the phosphoric oxide drying train (*C*, *D*, figure 1).

Hydrogen sulphide was prepared by direct union of hydrogen and sulphur in the vessel *E*. This tube, which was originally closed at *G* by a glass septum, was loaded with an excess of sulphur and evacuated through the connexion *F*. After again distilling the sulphur in a good vacuum to remove traces of air, sulphur oxides, etc., hydrogen (free from oxygen) was admitted to a pressure of about 550 mm., and the tube was sealed at *F*. It was then heated at 370° C for 20 hr., when, according to

Bodenstein's data (1899), practically complete conversion of hydrogen to hydrogen sulphide should have occurred; a small proportion of uncombined hydrogen was not significant for our purpose. The space above the septum *G* was then connected to the vacuum pump and evacuated, whereupon *G* was perforated by means of a glass breaker. This procedure was adopted as affording the simplest means of securing the necessary rigorous exclusion of oxygen from the hydrogen sulphide.

Resistance measurements were made by direct potentiometric determination of current and voltage drop across the specimen. The values found include, admittedly, any contact resistances, but, as stated above, we believe no serious error to have been introduced thereby. This procedure, rather than measurement of the voltage drop between probe electrodes, was adopted to avoid certain experimental difficulties. The validity of Ohm's law was confirmed by an independent series of measurements.

The general procedure with each specimen was as follows. The resistance-temperature characteristics of the pellet, as prepared and mounted in air, were first determined by an ascending and descending run with the apparatus evacuated to 10^{-3} – 10^{-4} mm. As our interest lay chiefly in the low-temperature phenomena, these runs were not usually carried out to temperatures high enough to intersect the intrinsic-conduction curve (see below). This omission we may remedy in a later series of measurements. Hydrogen was then introduced, and the temperature was raised to 300–360° C, when a rapid change in conductivity occurred. After a steady resistance had been reached and maintained for some hours, measurements between 350 and 18° C were made, with descending temperatures. A second hydrogen treatment, or a run in vacuum frequently followed. Restoration of conductivity ensued to a greater or less degree on admission of oxygen at the ordinary temperature (see tables), or of hydrogen sulphide, which reacted above 150° C. Such treatments were interspersed with reductions. At every temperature, adequate time was allowed to ensure that the measured resistance was that of an equilibrium state, and where there was any indication of a process producing a slow drift, the temperature was held steady for many hours. In several runs the specimen was cooled as quickly as possible from the highest temperature (air quench only), in order to freeze in lattice defects, if possible. With 'reduced' specimens at least, these experiments yielded negative results.

RESULTS

As has been found for other semiconductors, the variation of conductivity with temperature can be expressed by the relation

$$\kappa_T = A_1 e^{-E_1/kT} + A_2 e^{-E_2/kT}, \quad (1)$$

wherein the first term, for which A_1 , E_1 are relatively small ($A_1 = 1-10$; $E_1 = 0.1-0.35$ eV), dominates the conductivity at lower temperatures, while the second term ($A_2 = 300-1200$; $E_2 = ca. 0.6$ eV) determines the conductivity at higher temperatures.

From the general theory of semiconductors as developed by Wilson, it is seen that (cf. Mott & Gurney 1940) for the concentration of free electrons in a semiconductor with N potentially conducting centres in the volume V ,

$$\frac{n_f}{V} = \left(\frac{N}{V}\right)^{\frac{1}{2}} \left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}} e^{-E/2kT}, \quad (2)$$

where E is the energy of excitation of an electron from a filled band to the lowest empty level. As the exponential term in (2) is dominant, and as electronic mobilities should not be strongly temperature-dependent at elevated temperatures, (2) leads to the relation

$$\kappa_T = n_f eV = \kappa_0 e^{-E'/kT}. \quad (3)$$

For intrinsic conduction, N/V represents the number of atoms per unit volume, E' the energy gap between the uppermost filled levels and the conduction band. Owing to the magnitude of E' , the intrinsic conductivity of most semiconductors hitherto investigated has been negligible at ordinary temperatures. In the present instance we identify the high-temperature term in (1) with intrinsic conduction. The relevant data are listed in table 3, and are discussed in the last section of the paper. It is generally conceded that conduction at lower temperatures is attributable to the presence of impurity centres, the most important of which are furnished by deviations of semiconducting compounds from ideally stoichiometric composition. The concentration of electrons released from impurity centres is represented by an expression exactly analogous to (2), in which N/V is the concentration of impurity centres. Results for the conductivity in the lower temperature range (18–300°C), summarized for about 50 series of measurements by appropriate values for A_1 , E_1 and κ_{18° , are listed in table 1. It is evident that the low-temperature conductivity is markedly influenced by such chemical treatment as would be expected to modify a stoichiometric defect.

The results listed in table 1 may be summed up as follows:

(a) The conductivity of stannous sulphide was (as is the rule with semiconductors) variable from sample to sample, but was usually of the order $1\text{--}2 \times 10^{-2}$ ohm⁻¹ at 18°C.

(b) Treatment of the stannous sulphide with hydrogen at relatively low temperatures greatly diminished the conductivity. Thus, pellet III had an initial resistance of 289 ohms. Reduction with commercial hydrogen raised this [III. 2] to 29,000 ohms, and treatment with pure hydrogen to 560,000 ohms [III. 7]. Similar considerable changes in conductivity may be observed in every case (cf. runs IV. 2, V. 2, VI. 3, VIII. 2, etc.). At the temperatures employed (up to 350°C), reduction of stannous sulphide to metallic tin did not occur; the process reached a limit, shown by the attainment of a certain minimum conductivity (cf. III. 7, IV. 14, VI. 3) beyond which the removal of conducting centres did not proceed. These values of the minimum conductivity were of the same order of magnitude ($5\text{--}8 \times 10^{-6}$ ohm⁻¹) for different samples.

(c) On exposure to oxygen at the ordinary temperature, or to hydrogen sulphide at elevated temperatures, the conductivity was restored. Thus, in run IV. 18, the resistance of the pellet dropped almost instantaneously from 77,000 to 300 ohms on exposure to the atmosphere; in run VI. 1 the change was from 360,000 to 51 ohms; in run IX. 2, from 191,000 to 24 ohms. Inspection of table 1 will show that this behaviour was quite general, and also that the number of conducting centres introduced is some function of the partial pressure of oxygen or of hydrogen sulphide in the system. This may be illustrated by the results for pellet III. When the resistance had been raised to 29,000 ohms by reduction with commercial hydrogen [III. 2, III. 3], admission of fresh commercial hydrogen (representing a partial pressure of about 0.8 mm. of oxygen) reduced the resistance to 1130 ohms. Further reduction with pure hydrogen raised the resistance to 404,000 ohms [III. 5], but on opening the tube to the fore-pump vacuum (partial pressure of oxygen 0.01–0.05 mm.) the resistance again fell to 19,500 ohms [III. 6]. The original resistance was often closely reproduced when the samples were exposed to the air. The effect of admitting hydrogen sulphide [III. 9, IV. 5] was similar. The experiments in hydrogen-hydrogen sulphide mixtures were originally attempted in order to determine in what manner the conductivity was a function of the activity of sulphur in the gas phase. They were unsuccessful, however, as the steel of the press reacted with hydrogen sulphide at the higher temperatures. We not infrequently observed, therefore, that after an initial increase in conductivity, the stannous sulphide was again reduced by hydrogen, liberated through reaction with the press (cf. III. 10, VI. 4, etc.). In a number of runs (marked in table 1, column iii with a double asterisk), the conductivity diminished very markedly on heating to 300–400° C in a vacuum. We have some evidence that these represented normal reduction, brought about by a trace of hydrocarbon vapour from the joint grease or from a wax seal.

All these facts fall into place on the hypothesis that, in conformity with the Meyer-Friederich rule, stannous sulphide is an essentially non-stoichiometric compound, the crystal lattice being stable over a certain range of composition for which $\text{Sn}:\text{S} = 1:1+x$. The magnitude of the composition range is discussed in a later section. As in other compounds of this type, the excess of non-metal is attributable to the occurrence of a proportion of vacant tin lattice sites, and the conversion of a corresponding number of Sn^{2+} ions to Sn^{3+} or Sn^{4+} ions (less probably, conversion of S^{2-} ions to S^- ions, cf. De Boer & Verwey 1937) to maintain electrostatic balance. Every excess anion would thus introduce two positive holes, and stannous sulphide should be a positive hole conductor. Treatment with hydrogen diminishes the stoichiometric excess of non-metal, destroying positive holes and lowering the conductivity. Recovery of conductivity on exposure to oxygen or hydrogen sulphide represents the building of excess anions on to the lattice, with creation of new positive holes. It is evident that a stoichiometric excess of non-metal may be achieved by incorporation of either sulphur, proper to the lattice, or oxygen; also, that in the normal condition, i.e. exposed to air, stannous sulphide contains an excess of non-metal.

TABLE 1. SUMMARIZED DATA FOR LOW-TEMPERATURE CONDUCTION

run	treatment		κ_{18° ohm ⁻¹	E_1 eV	A_1
Pellet I. SnS sample 3, as prepared					
1	heated in vacuum	up	5.7×10^{-3}	0.15	2.2
2	heated in vacuum	up	6.2×10^{-3}	0.135	1.3
3	heated in vacuum	up	2.4×10^{-3}	0.14	0.6
4	heated in vacuum	up	7.1×10^{-3}	0.09	0.25
	after rapid cooling from 460° C		9.7×10^{-3}		
Pellet III. SnS sample 4b as prepared					
1	run in forepump vacuum	up	1.2×10^{-2}	0.15	4.5
2	commercial H ₂ (0.2 % O ₂) admitted		1.2×10^{-2}	—	—
	heated to 350° C	down	1.2×10^{-4}	0.213	0.6
3	readmission of commercial H ₂		3.0×10^{-3}	—	—
	heated to 300°	down	1.2×10^{-4}	see table 2	
4	on readmitting commercial H ₂		3.0×10^{-3}	—	—
	heated to 350°	down	4.5×10^{-5}	0.285	4.0
5	heated in H ₂ to 350° C	down	7.6×10^{-6}	0.345	6.8
6	exposed to forevacuum	up	1.7×10^{-4}	0.24	2.3
		down	9.6×10^{-6}	0.34	5.7
	admitted commercial H ₂		1.6×10^{-3}	—	—
7	heated in pure H ₂ to 200° C	down	6.2×10^{-6}	0.350	6.6
8	9 mm. H ₂ S admitted		8.2×10^{-5}	—	—
	heated to 200° C	down	1.5×10^{-5} *	0.310	4.0
9	29 mm. H ₂ S admitted	up	1.3×10^{-3}	0.205	4.5
	heated to 250° C	down	7.2×10^{-3}	0.160	4.4
10	60 mm. H ₂ S admitted	up	9.2×10^{-3}	0.155	4.8
	heated to 250° C	down	7.7×10^{-4} *	0.205	2.7
Pellet IV. SnS sample 4b as prepared					
1	heated in vacuum		1.1×10^{-2}	0.140	2.8
2	heated in H ₂ to 360° C	down	3.8×10^{-5}	0.27	1.5
3	fresh H ₂ admitted		3.8×10^{-5}	—	—
	heated to 400° C	down	2.4×10^{-5}	see table 2	
4	fresh H ₂ admitted		2.4×10^{-5}	0.280	1.8
	heated to 270° C	down	1.6×10^{-5}	0.310	3.4
5	33 mm. H ₂ S admitted		1.0×10^{-4}	—	—
	heated to 290° C	down	1.0×10^{-3}	0.140	2.8
6	26 mm. H ₂ S + 39 mm. H ₂				
	heated to 420° C	down	7.2×10^{-3}	0.145	2.5
	exposed to atmosphere		7.2×10^{-3}	—	—
10	heated in H ₂ to 380° C		'anomalous' curves. See table 2		
11	heated in 27 mm. H ₂ S + 116 mm. H ₂				
	to 400° C				
12	heated in vacuum to 425° C				
13	heated in vacuum to 400° C	up	2.7×10^{-2}	0.085	2.2
		down	2.7×10^{-2}	0.185	4.3
14	heated in H ₂ to 340° C	down	6.8×10^{-3}	see table 2	
15	heated in vacuum	up	—	0.36	10.2
	after chilling from 420° C		1.9×10^{-5}	—	—
16	heated in vacuum:	up		see table 2	
	after chilling from 470° C		2.2×10^{-5}		
	exposed to forevacuum		9.7×10^{-4}		
17	heated in vacuum to 360° C		1.5×10^{-5}	see table 2	
	exposed for forevacuum		1.3×10^{-3}		
18	heated in vacuum to 260° C		3.8×10^{-5}	see table 2	
	exposed to atmosphere		1.1×10^{-3}		
19	heated in vacuum to 400° C		3.3×10^{-5}	see table 2	

TABLE 1 (continued)

run	treatment		$\kappa_{18} \text{ ohm}^{-1}$	$E_1 \text{ eV}$	A_1
Pellet V. SnS sample 4b as prepared					
1	heated in vacuum to 300° C	up	2.0×10^{-3}	0.122	2.6
2	heated in hydrogen to 400° C	down	1.7×10^{-3}	see table 2	—
	exposed to atmosphere		1.4×10^{-3}	—	—
3	heated in vacuum to 450° C	down	$3.7 \times 10^{-3**}$	0.312	9.2
	admitted 27 mm. H_2S	"	3.3×10^{-4}	—	—
	exposed to atmosphere		3.0×10^{-3}	—	—
Pellet VI. SnS sample 4b after sublimation at ca. 700° C in high vacuum					
1	heated in vacuum to 350° C	up	8.5×10^{-3}	0.150	2.7
		down	$3.1 \times 10^{-3**}$	see table 2	—
	exposed to atmosphere		2.2×10^{-3}	—	—
2	heated in vacuum to 400° C		2.2×10^{-3}	0.120	2.5
3	heated in H_2 to 400° C	down	4.9×10^{-6}	see table 2	—
4	admitted 25 mm. H_2S	up	8.6×10^{-5}	0.260	2.6
		down	$6.7 \times 10^{-4**}$	see table 2	—
5	heated in vacuum to 215° C, then admitted H_2S , heated to 320° C exposed to atmosphere	down	$3.5 \times 10^{-5**}$ 5.3×10^{-2}	see table 2	—
Pellet VII. SnS sample 4b fused at 950° C in vacuum for 45 min.					
1	heated in vacuum to 230° C		2.8×10^{-3}	0.155	2.7
2	heated in H_2 to 410° C	down	1.3×10^{-4}	see table 2	—
3	26 mm. H_2S admitted at 235° C, then heated in vacuum to 300° C. Subsequently $\kappa_{18} = 1.1 \times 10^{-3}$		—	—	—
Pellet VIII. SnS sample 4c as prepared					
1	heated in vacuum to 275° C	up	3.5×10^{-2}	0.140	9.6
2	H_2 admitted at 275° C, then heated to 470° C		8.3×10^{-5}	see table 2	—
3	after prolonged and repeated treat- ment with H_2 up to 510° C exposed to 1 mm. air		2.2×10^{-3} 7.9×10^{-3}	see table 2	—
4	heated in vacuum to 300° C	up	7.9×10^{-3}	0.220	5.4
	subsequently exposed to atmosphere		$2.1 \times 10^{-4**}$ 2.1×10^{-1}	—	—
Pellet IX. SnS sample 4c fused at 950° C in vacuum for 8 hr. and then quenched					
1	heated in vacuum to 230° C	up	1.3×10^{-3}	0.155	6.7
2	H_2 admitted at 230° C, then heated to 380° C	down	5.7×10^{-6}	0.345	5.0
	exposed to atmosphere		4.5×10^{-2}	—	—
Pellet X. Material from SnS sample 4c fused at 950° C in vacuum for 8 hr., subsequently heated with H_2S in sealed tube at 800° C for 22 hr.					
1	heated in vacuum to 290° C	up	2.3×10^{-3}	0.105	1.65
2	H_2 admitted at 290° C, then heated to 390° C	down	2.0×10^{-5}	see table 2	—
	exposed to atmosphere		4.8×10^{-3}	—	—
3	heated in H_2 at 350° C for 48 hr.		9.5×10^{-5}	see table 2	—
4	heated in vacuum to 350° C		6.6×10^{-5}	see table 2	—

Footnote to table 1. Values marked *up* represent initial conductivity, κ_{18} , and E , A constants for series of measurements with ascending temperature. Values marked *down* similarly represent values obtained with descending temperature and subsequent value of κ_{18} .

* In these runs an initial increase in conductivity was observed through reaction with H_2S , but was nullified by reaction with the material of the press.

** In these runs a decrease in conductivity was observed without admission of hydrogen, attributed to reduction by hydrocarbon vapour.

The theory of defect lattices (cf. Schottky & Wagner 1930; Fowler & Guggenheim 1939) suggests that a strictly unilateral departure from stoichiometric composition is improbable, and that under appropriate conditions the existence of a tin-rich phase is to be expected, containing the excess of tin atoms in interstitial conditions and acting as an excess, or electron, conductor. This expectation was verified by experiments on pellet VIII, which was subjected to prolonged and repeated treatment with hydrogen at temperatures up to 510° C. Its conductivity passed through a minimum, and then rose as *additional* conducting centres were produced by the further abstraction of sulphur from the lattice. Values for the resistance of this pellet at 227° C, after successive hydrogen treatments, were 275, 245, 93 and 63 ohms. Following the last-mentioned stage there was some evidence of the production of metallic tin as a separate phase, and it may be inferred that the range of composition of the lattice on the tin-rich side is narrower than the permitted range of non-metal excess.

Pellets IX and X were prepared in order to examine the behaviour of material which, through the dissociation of the melt referred to previously, should have contained a slight initial excess of tin frozen into the lattice. The minimum conductivity of these samples was somewhat greater [IX. 2, X. 4] than that of the original material, whereas that of pellet VI (sublimed material, probably approximating most closely to ideal composition) was smaller. The general properties of the fused material however were not different from those of the original stannous sulphide—a fact that is not without bearing upon the distribution of positive holes, considered in a later paragraph.

The foregoing hypothesis was qualitatively substantiated by observation of the thermoelectric effect in the system Cu/SnS/Cu. This couple gave a large thermoelectromotive force, of about the magnitude computed from the activation energy of the conduction process, and indicative of the existence of *positive* current carriers in the stannous sulphide. Quantitative measurements of the effect, now in progress, will be described at a later date.

DISCUSSION

It will be seen from table 1 that whilst, for any sample, the conductivity at room temperature, κ_{18} , ran parallel with the variations in stoichiometric composition brought about by varied chemical treatment, the changes in conductivity arose almost entirely from the alteration in the activation energy E , since this enters as an exponent into equation (1), and not from changes in the factor A . E , in fact, must be a function of the concentration of impurity centres. Gudden (1934) has already pointed out that the relation between conductivity and activation energy is quantitatively the same for a variety of oxide semiconductors, and has justifiably queried the generally accepted model of an impurity conductor. In figure 2, the shaded area represents the locus of values of E and $\log \kappa_{18}$ within which fall practically all the recorded data for semiconductors, excepting a few points from the measurements

of Engelhard (1933) and Juse & Kurtschatow (1933). Even the very small and the negative values of E recorded, for example, by Meyer & Neldel, fall into place in figure 2. The relation between E and $\log \kappa_{18}$ is independent of the sign of the current carriers, not only as for different conductors (e.g. the positive hole conductor Cu_2O and the electron conductor ZnO), but also when a change-over of conduction mechanism takes place in one and the same conductor. Most of our experimental points refer to stannous sulphide with excess of non-metal, but the conductivities measured for pellet VIII, with excess tin, fall equally well within the shaded area.

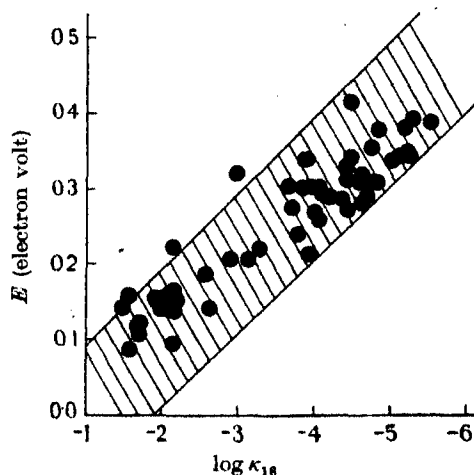


FIGURE 2

That E should be a function of the concentration of conducting centres has, at present, no satisfactory interpretation. It is not compatible with the concept of discrete impurity levels (cf. Schottky 1933) lying between the top of the full band and the conduction band, possessing energies characteristic of the particular impurity and host lattice concerned.

One attempt to explain the variability of E was made by Nijboer (1939), on the assumption that an excess of thermal lattice defects may be frozen in, so that the number of electrons n_b is less than the total number of impurity centres N . This leads to the relation

$$\frac{n_f(N - n_b + n_f)}{n_b - n_f} = Ae^{-E/kT}, \quad (4)$$

and, making plausible assumptions regarding the magnitude of N , n_b and E , Nijboer showed that (4) reproduced, with moderate success, the observed behaviour of alumina (Hartmann 1936). However, apart from the fact that (4) cannot give a linear relation between $\log n_f$ and $1/T$ when N and n_b are comparable (the important case for variability of apparent slope), it leads to the conclusion that the apparent activation energy, E_{app} , lies between the limits $E \geq E_{app} \geq \frac{1}{2}E$. This is in complete

disagreement with the measurements listed in table 1, as well as with the results of Fritsch and Meyer & Neldel. It may, perhaps, be held that Nijboer's theory suggests one reason for the occurrence of abnormal conductivity curves, exhibiting a discontinuity in slope. This point is discussed below.

Whereas E is evidently a function of the stoichiometric defect, the factor A (equation (1))—which, by hypothesis, depends upon both stoichiometric defect and electron mobility—does not vary greatly with any of our specimens (i.e. compared with the effect of the changes in E) and is, indeed, roughly the same for a variety of semiconductors. If the values of A (table 1) be examined, it will be evident that the lack of reproducibility inherent in the low-temperature process leads to a good deal of random scattering: no systematic trend of A is readily discernible. It has been considered by some (cf. Nijboer 1939; Seitz 1940) that A is independent of the stoichiometric defect. However, Meyer (1935), in a discussion of this point, concluded that A does depend upon the conductivity of the material. His data for titania, which can undergo very marked changes in composition (from TiO_2 to approximately $\text{TiO}_{1.75}$), with correspondingly great enhancement of conductivity, indicate that A becomes large for specimens with very high or very low conductivity, all the experimental points falling on one or other of two intersecting lines when $\log A$ is plotted against $\log \kappa_{20}$. The A values for stannous sulphide (including those for anomalous curves and for the high temperature conductivity), when subjected to a similar graphical analysis, show a distribution in general harmony with Meyer's conclusion, though the range covered is hardly sufficient to outweigh unambiguously the effect of scattering.

On the assumption that the electrons possess thermal velocities, and have the normal effective mass m , the mobility v is given by

$$v = \frac{el_0}{(3kTm)^{\frac{1}{2}}},$$

so that for the conductivity at temperature T

$$k_T = \frac{4\sqrt{2}}{3} n \frac{e^2 l_0}{h} (2mkT)^{\frac{1}{2}} e^{-E/2kT}.$$

Inserting numerical values,

$$A = 0.024 n \frac{1}{h} l_0 T^{\frac{1}{2}} \text{ ohm}^{-1} \text{ cm.}^{-1}. \quad (5)$$

Constancy of A would imply that $n \frac{1}{h} l_0$ is constant. In any case, the electronic mean free path must depend upon the concentration of impurity centres in a manner not readily interpreted. Such a relation finds no place in the present theory of semiconductors, e.g. in the discussion of the mean free path given by Mott & Frohlich (1939). It is evident that in this respect also the theory is in need of review.

In the absence of direct measurements of the stoichiometric defect we may reasonably examine the consequences of accepting the assumption (Mott & Frohlich 1939) that the mean free path is of the order of a few Ångström units, in order to calculate

n_b from (5). If, for example, $l_0 = 5 \text{ \AA}$, $n_b = 6 \times 10^{17}$. If the poorly conducting stannous sulphide be comparable with the poorly conducting cuprous oxide examined by Engelhard, for which $l_0 = 50 \text{ \AA}$, $n_b = 6 \times 10^{15}$. These figures may be taken as reasonable indications of the range of stoichiometric defect needed to account for the variation in conductivity observed in our experiments. The compressed stannous sulphide contained about 1.6×10^{22} atoms of tin and of sulphur per c.c., so that the range of defect would correspond with perhaps 10^{-4} to 10^{-7} atom excess sulphur per atom of tin. To determine these quantities directly, measurements of the numbers of free electrons, from the Hall coefficients, would be needed.

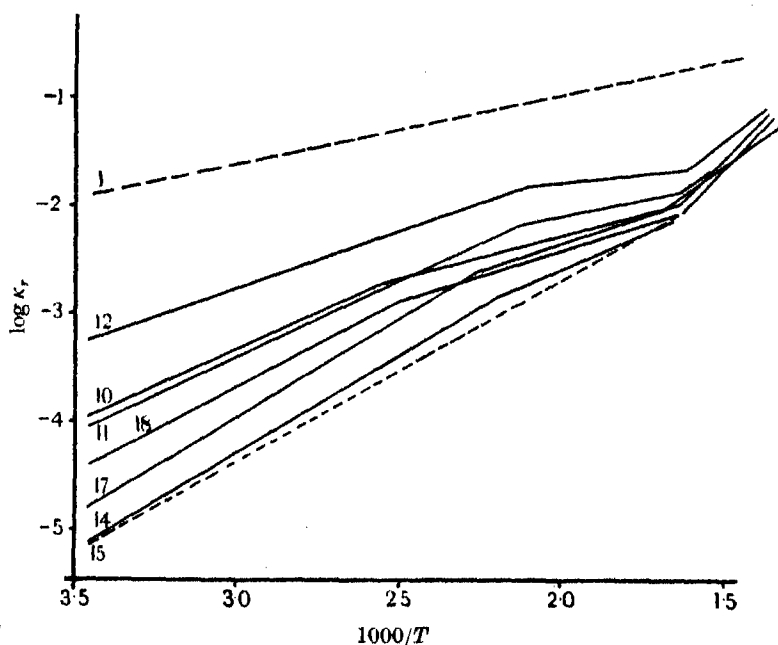


FIGURE 3

Abnormal conductivity-temperature curves. Reference has already been made to conductivity-temperature curves exhibiting a kink at a certain temperature, above which the relation between $\log \kappa_T$ and $1/T$ was again linear, but with a different, abnormally small slope. Discontinuities of this kind in the characteristic constants defining the conductivity may be seen also in the results of Vogt & Engelhard for cuprous oxide, and possibly in those of Jander & Stamm (1931), von Baumbach & Wagner (1933) and Miller (1941) for zinc oxide, though we incline to the view that the zinc oxide data are complicated by changes in the stoichiometric defect. Figure 3, representing results obtained on pellet IV, shows the characteristics of these abnormal curves; the principal results are collected in table 2, and the previous and subsequent history of the samples may be obtained from table 1.

TABLE 2. CHARACTERISTIC DATA FOR ABNORMAL CONDUCTIVITY CURVES

run	treatment	κ_{18}	κ_{27}	E_1	A_1	E_A	A_A	T_e
III. 3	heated in commercial H_2 to 300° C	down						
IV. 3	heated in pure H_2 to 400° C	down	7.4×10^{-3}	0.304	21.4	0.17	0.4	115°
IV. 10	heated in pure H_2 to 380° C	down	3.5×10^{-3}	0.320	8.1	0.239	0.9	160°
IV. 11	heated in 27 mm. H_2S + 116 mm. H_2 to 400° C	down	5.1×10^{-3}	0.274	10.8	0.137	0.1	120°
			8.7×10^{-3}	0.305	1616	0.153	0.3	180°
IV. 12	heated in vacuum to 425° C	up	1.2×10^{-2}	0.220	3.5	0.098	0.1	200°
IV. 14	heated in H_2 to 340° C	down	2.7×10^{-3}	0.384	28.8	0.261	1.1	171°
IV. 16	heated in vacuum	up	4.5×10^{-3}	0.310	5.9	0.223	0.8	227°
IV. 17	heated in vacuum to 360° C	down	4.9×10^{-3}	0.380	54.9	0.207	0.6	158°
IV. 18	heated in vacuum to 260° C	down	3.8×10^{-3}	0.330	19.0	0.199	0.4	125°
IV. 19	heated in vacuum to 400° C	down	1.2×10^{-2}	0.418	549	0.171	not linear	
V. 1	reduction (hydrocarbon vapour?)	down	2.3×10^{-2}	0.303	40.3		1.2	170°
	on heating in vacuum							
V. 2	heated in H_2 to 400° C	down	5.9×10^{-3}	0.356	24	—	—	—
VI. 1	reduction (hydrocarbon vapour?)	down	2.5×10^{-3}	0.438	115	—	—	—
	on heating in vacuum							
VI. 3	heated in H_2 to 400° C	down	2.5×10^{-3}	0.385	21.9	0.300	26	192°
VI. 4	heated in vacuum after H_2S treatment	down	2.4×10^{-3}	0.374	19.5	0.237	0.6	176°
VI. 5	evacuated after treatment with H_2S at 215–285° C	down	5.6×10^{-3}	0.344	219	0.05	0.02	192°
VII. 1	evacuated after H_2S treatment	down	1.1×10^{-2}	0.342	105	0.006	0.01	145°
VII. 2	heated in H_2 to 410° C	down	6.1×10^{-2}	0.318	330	—	not linear	
VIII. 2	heated in H_2 to 470° C	down	1.0×10^{-3}	0.292	—	0.167	—	—
VIII. 3a	heated in H_2 to 320° C		7.5×10^{-3}	0.264	3.4	0.143	0.2	219°
VIII. 3b	fresh H_2 ; 18 hr. at 330° C		8.3×10^{-3}	—	—	0.147	0.25	—
VIII. 3c	fresh H_2 ; 45 hr. at 370° C		1.4×10^{-4}	0.302	23.3	0.062	0.09	227°
VIII. 3d	fresh H_2 ; 4 hr. up to 510° C		—	—	—	0.102	0.35	215°
VIII. 3e	reheated to 360° C	down	3.3×10^{-3}	0.342	112	0.102	0.35	215°
X. 2	heated in H_2 to 390° C	down	1.4×10^{-4}	0.292	2.2	—	not linear	
X. 3	evacuated after H_2 treatment	down	2.0×10^{-5}	0.268	4.1	—	not linear	
X. 4	heated in vacuum to 350° C	down	9.5×10^{-5}	0.288	6.3	0.04	0.005	140°
			6.6×10^{-5}					

Abnormal curves were obtained only for poorly conducting material, i.e. when the number of vacant cation sites determined by the stoichiometric defect was relatively small. For such samples the occurrence of kinked curves, though by no means invariable, was common, especially after the pellet had been subjected to repeated treatment with hydrogen and oxygen or hydrogen sulphide; that is, when by repeated addition of anions to the lattice and removal of anions from it, a high degree of imperfection could have been created (cf. Nijboer). The lowest recorded conductivities were obtained in runs III. 5, III. 7, IV. 15, IX. 2; these were normal curves, although the same pellets in previous and subsequent runs displayed the abnormal behaviour.

From table 2 it will be seen that the activation energy for the low-temperature arm of the curve, E_l , has values comparable with those of normal curves, although rather larger than would be expected from the observed values of κ_{18} . The activation energy of the high-temperature segment, E_h , is frequently round about $0.4-0.6E_l$, but may be very small. It is characteristic of the abnormal curves that A_l is higher, by a factor of about 10, than for normal curves, while A_h is smaller by a similar factor. These features might be attributable to changes in the mean free path (due, for example, to high concentrations of lattice imperfections), or in the effective number of positive holes, or to both. Two purely empirical relations may be discerned: (a) A for these curves is a function of E , $\log A_l$, $\log A_h$ being roughly proportional to E_l , E_h respectively; (b) for all the curves, $\log A_l/A_h$ is closely proportional to $E_l - E_h$. These relations may be fortuitous. The temperature T_c at which the discontinuity in the A , E values occurred showed a general downward trend with decrease in stoichiometric defect. Kinks were observed with both ascending and descending temperatures, but whereas with normal curves the ascending and descending measurements made on the same sample without change in composition were closely coincident, with the abnormal curves successive runs showed a marked lack of reproducibility of A_l , E_l , A_h , E_h , even though the conductivities at 18°C and at the temperature of onset of the 'high-temperature' conduction were the same. It was carefully established that there was no drift in the measurements. Points on both the upper and the lower segments of the curves refer to steady states of the material.

We abstain from offering any explanation for these curves. They cannot be assigned (cf. Engelhard) to phase transitions; Nijboer's hypothesis is excluded by the range covered by the ratio $E_l:E_h$; Miller's view, that they represent the onset of saturation effects, is not in harmony with his own experimental data, and incompatible with the usual linearity of the curves above T_c .

Production and distribution of lattice defects. There is one other aspect of our measurements that appears worthy of comment. This is the rate at which the chemical processes concerned in the change in stoichiometric defect occur. When stannous sulphide, in equilibrium with air, was treated with hydrogen, no measurable reaction took place below about 200°C , but at 300°C the reduction process evidenced by diminished conductivity proceeded with great rapidity. On exposure

of poorly conducting stannous sulphide to air (or to hydrogen sulphide above 150–180° C) the fall of resistance was, in its early stages, immeasurably rapid. Thus, at the conclusion of run VIII. 4 (cf. table 1), the following observations were made:

Resistance initially	9500 ohms
Less than 30 sec. after exposure to air (1st measurement)	100 ohms
After 90 sec.	70 ohms
After 5 min.	30 ohms
After 60 min.	15 ohms
Final resistance	10 ohms

Reactions within solids can occur only through the migration of atoms or ions. The energy of activation for diffusion in ionic lattices is of the order of 1 eV; for diffusion of atoms in the lattice of metallic sulphides, where the forces are probably largely covalent, the activation energy may well be higher. Hence, such reactions cannot occur at the ordinary temperature, and we may conclude that the reactions, whereby positive holes were created or destroyed, occurred exclusively at the surface of the crystals.

It is of interest to examine how far the probable surface area of a crystalline powder is adequate to account for the estimated concentration of lattice defects. If the size of crystallites, or of blocks in the crystal mosaic, were 10^{-3} mm., or about 2×10^3 times the averaged unit cell dimension of stannous sulphide, in 1 c.c. of the material about 2.5×10^{19} atoms of Sn or S would be present in the surfaces; larger or smaller crystallite sizes would give proportionate changes in the number of surface atoms. It is evident that the number of atoms exposed in the surfaces of a crystalline material may well be sufficient to allow of changes of composition of the order estimated, without affecting the underlying lattice at all. At high temperatures, where diffusion can occur, an equilibrium distribution of defects between surface and interior would ultimately result; positive holes in the interior of the lattice would be withdrawn from ready reaction. It follows, further, that the estimate made for the concentration of impurity centres may represent only a fraction of the total stoichiometric defect.

In this connexion it is worth considering the results obtained with sublimed and with fused stannous sulphide. The material of pellet VI, formed by condensation of SnS molecules from the vapour phase, should have approached very closely to the ideal stoichiometric formula. Handled in air, reaction occurred at the surface, with the creation of positive holes, as with other specimens, but the maximum conductivity produced thereby, and the minimum conductivity after reduction with hydrogen were similar to those found for other specimens with very different thermal and chemical histories. Furthermore, the minimum conductivities recorded for fused material (pellets IX, X)—which might be expected to contain a high proportion of frozen-in lattice defects—were not markedly different from those of the standard or the sublimed material. The conclusion seems inevitable that surface

effects are of great, and even preponderating importance in the low-temperature conduction process. It is probable that these considerations are equally cogent for many of the results obtained with other semiconductors.

The high-temperature conductivity

As has been found for other semiconductors investigated over a sufficiently wide range of temperature, the conductivities at high temperatures fell on a second rectilinear arm of the $\log \kappa - 1/T$ curve, characterized by higher values of the constants A and E (equation (1)). The principal data for this region are summarized in table 3.

TABLE 3. SUMMARIZED DATA FOR HIGH-TEMPERATURE CONDUCTION

run	κ_{18}	κ_{294}	E_2	A_2
IV. 3	2.4×10^{-5}	2.5×10^{-2}	0.62	1320
IV. 7	7.2×10^{-3}	3.9×10^{-1}	0.48	1190
IV. 11	9.0×10^{-4}	2.8×10^{-2}	0.52	233
IV. 12	5.5×10^{-4}	4.2×10^{-2}	0.47	253
IV. 15	6.8×10^{-6}	2.4×10^{-2}	0.61	1080
IV. 17	1.5×10^{-5}	2.8×10^{-2}	0.64	1830
V. 2	1.7×10^{-5}	3.4×10^{-2}	0.60	1130
V. 3	3.5×10^{-5}	3.4×10^{-2}	0.53	346
VI. 3	4.9×10^{-6}	2.8×10^{-2}	0.60	920
VI. 4	6.7×10^{-6}	2.8×10^{-2}	0.59	890
VI. 5	3.5×10^{-5}	3.1×10^{-2}	0.61	1170
VII. 2	1.3×10^{-4}	2.6×10^{-2}	0.55	390
VIII. 3e	1.4×10^{-4}	4.4×10^{-2}	0.56	960
X. 2	2.0×10^{-3}	1.2×10^{-2}	0.59	340
X. 3	9.5×10^{-5}	1.6×10^{-2}	0.63	895
X. 4	6.6×10^{-5}	1.6×10^{-2}	0.64	*1030

As the main concern of this work was with the low-temperature conductivity, relatively less attention was devoted to the high-temperature measurements. In a number of instances the measurements were made over a comparatively small range of temperature, and the values of table 3 represent what appear the most reliable data. With a few exceptions there is a general accord in the magnitude of E_2 ; some scattering is probably due, in part, to measurement of the slope near the junction of the high-temperature and low-temperature curves. Two series of measurements, not recorded, carried up beyond the melting-point, showed that there was no further change in slope of the conductivity curve at higher temperatures.

The high-temperature conductivity can reasonably be regarded as the intrinsic conductivity of the lattice, corresponding to a conduction band 1.2 V above the full band. This view is consistent with the characteristic constants of the process. The A values are highly sensitive to variations in E , but from table 3, A_2 may be taken as 200–1100. Since there are about 1.6×10^{22} atoms of tin per c.c. of the compressed pellets, then by substitution in equation (5) a value of 1.4–7 Å for the mean free path is obtained, or 3–19 cm. sec.⁻¹ cm.⁻¹ sec.⁻¹ for the mobility. These values are of the right order of magnitude.

The absolute values of the conductivity of different specimens, irrespective of chemical and thermal history, are in moderately close accord, as shown by the values of κ_{394° . By extrapolation, the intrinsic conductivity at 18° is about 10^{-7} – 10^{-8} , a value considerably in excess of conductivities usually found for ionic compounds (e.g. oxides with inert-gas-like cations), but in keeping with the considerations raised in the introduction.

REFERENCES

- Anderson & Ridge 1943 *Trans. Faraday Soc.* **39**, 98.
v. Baumbach & Wagner 1933 *Z. phys. Chem. B*, **22**, 109.
Bodenstein 1899 *Z. phys. Chem.* **29**, 315.
de Boer & Verwey 1937 *Proc. Phys. Soc.* **49**, 59.
Engelhard 1933 *Ann. Phys., Lpz.*, (5) **17**, 501.
Fischbeck & Dörner 1929 *Z. anorg. Chem.* **181**, 372.
Frey 1930 *Z. Elektrochem.* **36**, 511.
Friederich 1925 *Z. Phys.* **31**, 813.
Fritsch 1935 *Ann. Phys., Lpz.*, (5), **22**, 375.
Fowler & Guggenheim 1939 *Statistical Thermodynamics*, pp. 542–552. Cambridge University Press.
Gudden 1934 *Ergebn. exact. Naturw.* **13**, 222.
Guinchant 1902 *C.R. Acad. Sci., Paris*, **134**, 1224.
Hartmann 1936 *Z. Phys.* **102**, 709.
Hintenberger 1942 *Z. Phys.* **119**, 1. *Chem. Abstr.* **36**, 6391.
Hittorf 1851 *Ann. Phys., Lpz.*, **84**, 20.
Jander & Stamm 1931 *Z. anorg. Chem.* **199**, 165.
Juse & Kurtschatow 1933 *Phys. Z. Sowjet.* **2**, 453.
Koenigsberger 1910 *Ann. Phys., Lpz.*, (4) **32**, 179.
Koenigsberger 1914 *Jb. Radioakt.* **11**, 84.
Mark 1940 *Physical chemistry of high polymeric systems*, Section D. New York: Interscience Publ. Co.
Meyer 1933 *Z. Phys.* **85**, 278.
Meyer 1935 *Phys. Z.* **36**, 749.
Meyer & Neldel 1937 *Phys. Z.* **38**, 1014.
Miller 1941 *Phys. Rev.* **60**, 890.
Mott & Gurney 1940 *Electronic processes in ionic crystals*, chapter 5. Oxford.
Mott & Frohlich 1939 *Proc. Roy. Soc. A*, **171**, 496.
Nijboer 1939 *Proc. Phys. Soc.* **51**, 575.
Schottky 1933 *Phys. Z.* **34**, 858.
Schottky & Wagner 1930 *Z. phys. Chem. B*, **11**, 163.
Seitz 1940 *Modern theory of solids*, p. 192. McGraw-Hill.
Streintz 1902 *Ann. Phys., Lpz.* (4), **9**, 854.
Trey 1925 *Phys. Z.* **26**, 849.
Tubandt & Haedicke 1927 *Z. anorg. Chem.* **160**, 297.

Reversible adsorption of proteins at the oil/water interface

I. Preferential adsorption of proteins at charged oil/water interfaces

BY J. J. ELKES,* A. C. FRAZER,* J. H. SCHULMAN AND H. C. STEWART*

Pharmacology Department, University of Birmingham; Colloid Science Department, Cambridge; and Physiology Department, St Mary's Hospital Medical School, London

(Communicated by E. K. Rideal, F.R.S.—Received 13 September 1944)

[Plates 4–5]

The behaviour of positively and negatively charged oil-in-water emulsions, stabilized with hexadecyl trimethyl ammonium bromide and sodium hexadecyl sulphate respectively in the presence of protein solutions has been studied.

Under certain conditions proteins will adsorb to a charged oil/water interface. When finely dispersed oil-in-water emulsion was used to provide this oil/water interface, adsorption of protein resulted in flocculation of the oil droplets.

Flocculation of emulsion on the addition of protein is pH conditioned and occurred on the acid side of the isoelectric point of the protein with negatively charged and on the alkaline side with positively charged oil globules. No flocculation occurred on the alkaline side of the isoelectric point with a negative emulsion or the acid side with a positive emulsion.

The amount of protein required to cause maximum clarification of the subnatant fluid corresponded with that needed to give a firmly gelled protein monolayer at the interface, namely, 2.5 mg. of protein/sq.m. of interfacial area. With that amount of protein the flocculated oil globules remained discrete and no coalescence or liberation of free oil occurred. If only 1 mg. of protein/sq.m. of interfacial area was added, flocculation was followed by rapid coalescence of oil globules and liberation of free oil. If smaller amounts still were used, no visible change in the dispersion of the oil droplets could be seen macroscopically. With greater amounts than 2.5 mg./sq.m. of interfacial area, up to ten times the monolayer concentration was adsorbed to the interface.

Sodium chloride affected the flocculation range, and instead of the clear-cut change-over between the positive and negative interfaces at the isoelectric point of the protein, overlapping occurred. 5 % sodium chloride shifted the flocculation point about 1 unit of pH. The addition of sodium chloride also altered the point of maximum clarification. Thus with haemoglobin the maximum clarification point was shifted from 2.5 to 1.7 mg./sq.m. of interfacial area by the addition of 1 % sodium chloride.

The adsorption of protein on to charged oil/water interfaces was reversible. This was best demonstrated with haemoglobin. Thus, haemoglobin was adsorbed at pH 5.0 to a negative emulsion—the red floccules were washed and transferred to a buffer at pH 10. The haemoglobin was released and the emulsion was redispersed.

The effect of adsorption and desorption on the structure of the protein molecule has been studied with haemoglobin. By solubility and colour tests it was shown that the haemoglobin molecule was changed to para-haematin by adsorption and subsequent desorption from a charged oil/water interface.

Molecular weight and shape determinations were carried out on the desorbed protein.

Two proteins have been separated by this adsorption mechanism. This was demonstrated on a mixture of albumin and haemoglobin.

Some applications of the flocculation technique are indicated and the significance of the phenomena described are discussed.

* Sir Halley Stewart Research Fellows.

Recent work on the significance of emulsification of fat in intestinal absorption (Frazer 1943; Frazer, Schulman & Stewart 1944), the structure of the chylomicron in the blood (Elkes, Frazer & Stewart 1939), and the reversible detoxication of snake venom and bacterial toxins by oil-in-water emulsions (Frazer & Walsh 1934, 1939; Frazer & Stewart 1940) has encouraged a quantitative approach to the study of the charged oil/water interface in relation to proteins by the use of surface physico-chemical methods.

The behaviour of protein molecules when they are brought into relation with an air/water or oil/water interface has been extensively studied. The protein molecule has three main possible configurations dependent upon its position in the water phase before or after adsorption or when it is fixed at the interface. The effects of these changes in structure on the biological activity of the protein molecule is of importance. If the oil/water interface is in the form of a finely dispersed oil-in-water emulsion, the adsorption of the protein to this interface results in changes in dispersion and stability of the emulsion. Changes of this nature in the particulate fat in the body may have biological significance. The object of this paper is to put forward evidence of the pH conditioning and the reversible nature of adsorption at the charged oil/water interface in oil-in-water emulsions. Evidence of structural changes in the haemoglobin molecule after adsorption is given and the preferential adsorption of one protein from a mixture of two proteins is described. The possible significance of these observations is discussed.

TECHNIQUE

1. *Emulsions and interfaces*

The emulsions used throughout the experiments described were 5 % olive oil or paraffin emulsions, except where otherwise stated. 0.2 % of stabilizer was used in these emulsions, and they were prepared using the apparatus described by Frazer & Walsh (1933). The particle size was of the order of 0.5μ which gives an oil/water interfacial area of approximately 0.6 sq.m./c.c. of emulsion.

The stabilizer used to give a negatively charged interface was sodium hexadecyl sulphate, and hexadecyl trimethyl ammonium bromide was used to give a positively charged interface. The 0.2 % concentration of stabilizer was chosen since most of the agent would thus be at the interface and not in the continuous phase. It further ensures a saturated interfacial film of the stabilizer.

2. *Protein solutions*

Solutions of the serum proteins were prepared by fractional ammonium sulphate precipitation and dialysis against distilled water; 0.6 % NaCl was used for γ globulin. The albumin fraction was recrystallized by the method of Adair & Robinson (1930). The oxyhaemoglobin was prepared by lysis of washed cells with distilled water, centrifugation and subsequent dialysis against distilled water. A small number of experiments have been made using carboxyhaemoglobin. The concentration of the

protein solutions was determined by Adair's refractometric method and by micro-Kjeldahl. The haemoglobin concentration was determined by colorimetry and protein estimation. The structure of the molecule was studied by spectroscopy, determination of osmotic pressure and molecular weight, sedimentation constant, solubility and colour tests as described in the text.

The serum proteins used for most experiments were from the horse, and the fractions were derived from Burroughs and Wellcome's no. 3 horse serum. A small number of experiments were conducted with human protein fractions. The haemoglobin used was from human red cells and a series has also been studied using sheep haemoglobin.

3. *Buffer solutions*

The buffer range studied was pH 3.5–10.0. The buffers used had the following composition, and were all adjusted to a concentration sufficient to buffer against strong protein solutions. The pH of all buffers and mixtures was checked by glass electrode measurements:

Range pH 3.5–4.8, Sørensen's citrate/HCl.

Range pH 4.8–5.5, Sørensen's citrate/NaOH.

Range pH 5.8–8.6, dihydrogen phosphate/NaOH.

Range pH 8.0–10.0, Boric acid/NaOH.

These buffers were compiled from the formulae given by Clark (1928).

Di- and triphosphates in acid solution were avoided in the buffer range, since the polyvalent anions would discharge the positive interface formed by hexadecyl trimethyl ammonium bromide and break these emulsions.

4. *Flocculation method and criteria*

The buffer and protein solutions were mixed, and to this was immediately added the emulsion. After mixture, the tubes were kept at room temperature and examined macroscopically with the hand lens and microscopically with dark-ground illumination. Observations were repeated every hour for 3 hr. and finally after 24 hr., which allowed for maximum clarification in flocculated specimens. The following phenomena were observed:

(a) *No change*: the whole tube appeared as a homogeneous milky fluid. Under dark-ground the particles were seen to be single and discrete and in violent Brownian movement.

(b) *Flocculation*: the particles aggregated in clumps and these could be seen macroscopically like pepper grains on the side of the tube if it was tilted. Later these clumps 'lifted' and formed a thick white layer with no free oil at the top of the tube while the supernatant fluid clarified and became water clear. Under dark-ground these particles were seen in clumps but the individual particle size did not change. There was no Brownian movement (figure 2B, plate 4).

(c) *Breaking of the emulsion*: the particles ran together and formed a cream on the top of the fluid. Gradually free oil was seen on the side of the tube and finally a layer

of free oil separated at the top. Under dark-ground, coalescence and an increase in size of the oil particles could be seen. There was no Brownian movement. With coloured proteins such as haemoglobin a further aid to interpretation was afforded by the transference of the pigment with the protein from the continuous phase to the interface or vice versa.

EXPERIMENTAL RESULTS

1. Flocculation over the pH range 4.0–10.0 with positively and negatively charged interfaces

Each series was put up using equal quantities of buffer and protein solution; to this mixture a further aliquot of emulsion was added. The changes in the dispersion of the emulsion were noted, and in flocculated tubes maximum clarification of the

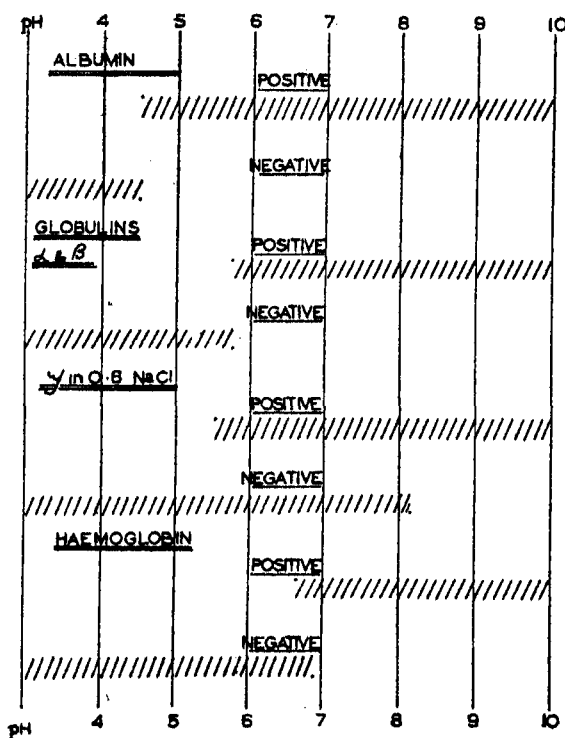


FIGURE 1. Flocculation reactions of proteins with positively and negatively charged oil-in-water emulsions. Cross-hatching indicates the flocculation range in each case.

subnatant fluid was obtained in 24 hr. The changes appeared immediately and became more pronounced overnight. The results of these experiments with albumin α , β , and γ globulin and haemoglobin are shown in figure 1 and figure 3 A, B, plate 4. Probable isoelectric points for the proteins are respectively pH 4.65, 5.2, 6.4 and 6.9 as determined by other methods (Svedberg & Pedersen 1940). The isoelectric points

determined by this method show identical figures for serum albumin and haemoglobin, but are somewhat higher for the globulins, being 5.8 and 6.8 respectively. It can be seen that flocculation occurred with each protein on the acid side of its isoelectric point if a negatively charged interface was used, and on the alkaline side with a positively charged interface. No change was found on the alkaline and acid sides of the isoelectric point with negatively and positively charged surfaces respectively within the stability ranges of these proteins.

2. The effect of protein concentration and interfacial area on the flocculation phenomenon

Numerous series of mixtures were studied with variable protein content and constant interfacial area and using constant protein with varying emulsion concentration.

(a) Point of maximum clarification threshold (table 1)

The point of maximum clarification of the supernatant fluid was found to be when 2.5 mg. of protein was added per sq.m. of interfacial area. Identical results were obtained with both negative and positive interfaces at appropriate pH levels to give flocculation and also in experiments in which the protein was kept constant and the interfacial area was varied. At threshold or greater concentration of protein there was no breaking of the emulsion.

TABLE 1

Determination of point of maximum clarification. pH 5.0.

Variable protein: constant interfacial area.

Protein: human haemoglobin.

Interface: 50 % paraffin-in-water emulsion stabilized with 2 % sodium hexadecyl sulphate.

In each tube 1 c.c. of dilute protein solution + 0.1 c.c. of emulsion.

Flocculation occurs throughout.

tube	...	1	2	3	4	5	6	7	8	9	10	11	12
Hb conc.	...	$\frac{1}{100}$	$\frac{1}{200}$	$\frac{1}{300}$	$\frac{1}{400}$	$\frac{1}{500}$	$\frac{1}{600}$	$\frac{1}{700}$	$\frac{1}{800}$	$\frac{1}{900}$	$\frac{1}{1000}$	$\frac{1}{1500}$	$\frac{1}{2000}$
subnatant fluid		← increasing depth of red colour due to free Hb					clear, no free Hb		increasing turbidity due to free emulsion →				

If 1 % sodium chloride added to each tube

← increasing depth of red colour due to free Hb					clear, no free Hb		increasing turbidity due to free emulsion →
---	--	--	--	--	-------------------	--	---

(b) Subthreshold concentrations (table 2 and figure 4, plate 5)

When a subthreshold concentration was added, two effects were seen. With grossly subthreshold amounts no visual effect was obtained. With a just-subthreshold level of about 1 mg./sq.m. interfacial area, the emulsion broke and free oil accumulated

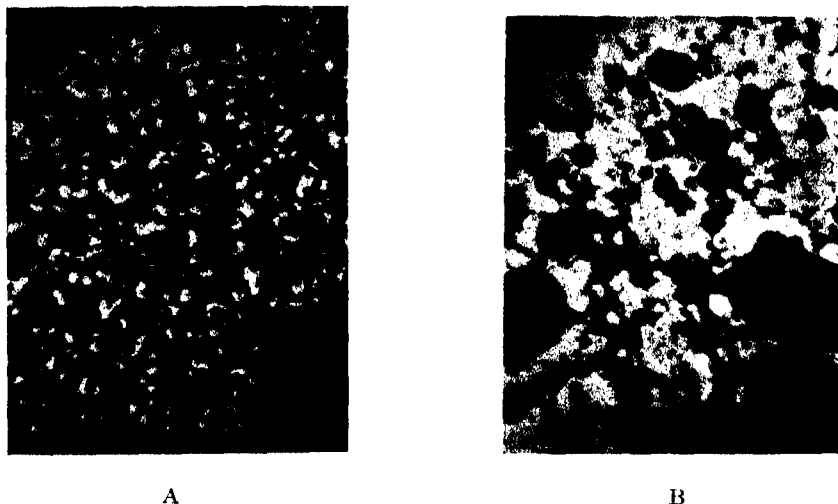


FIGURE 2. Untouched micro-photographs to show the difference between flocculation and breaking of emulsion. The photographs were taken of specimens under dark-ground illumination using 1/6 in. objective and $\times 8$ eyepiece (mag. $\times 96$), 2 min. exposure. The emulsion before flocculation showed a mass of free particles in violent Brownian movement. The individual particles were only just visible under this magnification. A. Flocculation. Particles in clumps; no Brownian movement; no free oil liberated. B. Breaking. Particles in clumps; no Brownian movement; large masses of free oil separating out.

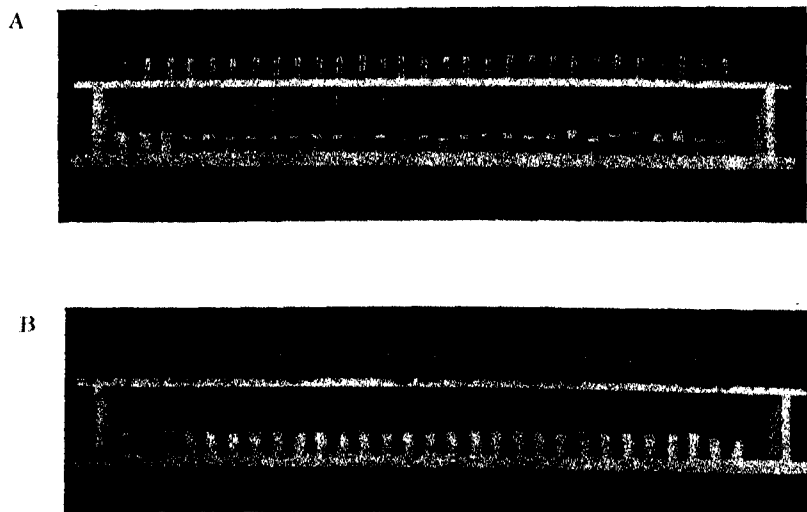


FIGURE 3. Untouched photograph of characteristic flocculation range experiment. Both series of tubes contain equal amounts of protein solution and emulsion with buffer added, so that the resultant pH ranges from 4 on the left of the rack to 10 on the right-hand side. In series A a positively charged emulsion was used, and in series B a negatively charged emulsion. It will be seen that flocculation occurs on the alkaline side of the isoelectric point in A and to the acid side in B.

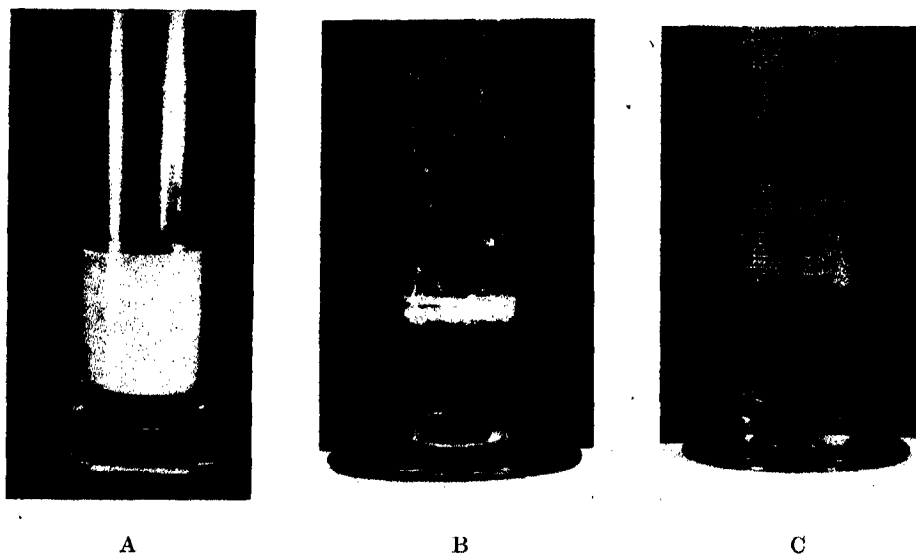


FIGURE 4. Untouched photograph of three mixtures of protein and emulsion with variable protein content. A. Emulsion to which a grossly subthreshold amount of protein has been added. There is no flocculation visible macroscopically. B. Emulsion to which a protein concentration of about 1 mg./sq.m. of interfacial area has been added. Flocculation and coalescence of the fat globules has occurred. Free oil can be seen on the surface. C. Emulsion to which 2.5 mg./sq.m. of interfacial area of protein has been added. Flocculation has occurred and the flocculated oil globules have remained discreet. No free oil can be seen in this specimen.

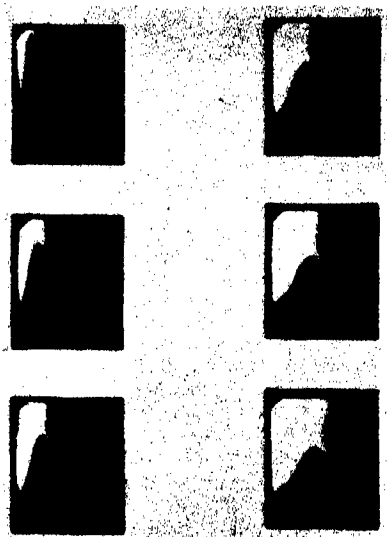


FIGURE 5. A solution of desorbed haemoglobin removed from a negative emulsion at pH 10 and ultracentrifuged. It was adsorbed at pH 5.0. Six 'Schlieren' photographs are given in course of the centrifuging:

	$S_{20}^{\circ} \times 10^{13}$	f/f_0
Haemoglobin	4.48	1.16
Desorbed haemoglobin (Parahaematin)	3.9	1.33

Phosphate buffer at pH = 8.6

$$\eta_{\text{rel.}} = \frac{\eta_{\text{buffer}}}{\eta_{\text{water}}} = 1.020 \pm (0.005).$$

Concentration of protein = 0.50 g./100 c.c.:

Rev./sec. ≈ 1266

Centrifugal field $\approx 182,000$ g.

In addition to the peak which gives a measure of dc/dx at different positions of the cell, owing to the colour of the protein the depth of the shading in the background makes possible an estimate of protein concentration.

at the surface. The normal emulsion was resistant to this same pH. The breaking can be accelerated if the flocculated emulsion particles are resuspended in fresh buffer at the same pH.

TABLE 2

Flocculation at threshold level.
 Constant protein, variable interfacial area. pH 4.6.
 Protein: sheep's haemoglobin.
 Interface: 5 % olive oil emulsion, stabilized with 0.2 % sodium hexadecyl sulphate.
 0.9 c.c. of a 1/15,000 haemoglobin solution in each tube.
 Emulsion diluted progressively as indicated, and 0.1 c.c. of dilutions added to corresponding tubes.

Flocculation occurs throughout.						
tube	...	1	2	3	4	5
dilution of emulsion		1/1	1/2	1/4	1/8	1/16
free oil		++	none	none	none	none
subnatant fluid		many free emulsion particles	no free emulsion particles	no free emulsion particles	no free emulsion particles	no free emulsion particles

(c) *Super-threshold concentrations*

Using haemoglobin, the effect of super-threshold concentrations was studied:

(i) When a constant amount of emulsion was added to varying concentrations of haemoglobin only a threshold or possibly a double threshold concentration allowed of complete clearing of haemoglobin from the subnatant fluid. Owing to the decrease in the interfacial area by flocculation, greater amounts of haemoglobin were not adsorbed.

(ii) When a 5 % emulsion was added to a strong haemoglobin solution (2 %), adsorption of greater amounts of haemoglobin up to approximately 20 mg./sq.m. of interfacial area occurred. This might be due to the greater area available before flocculation could be effective.

3. *Effect of sodium chloride concentration*

Two effects have been observed:

(a) Increasing concentration of NaCl displaced the flocculation range towards the alkaline side with a negative interface and to the acid side with a positive interface causing an overlap of the two flocculation ranges. This phenomenon can be easily demonstrated with albumin or γ globulin, but only slight overlapping occurs with α and β globulins. A gradation of overlapping is obtained with increasing concentration of sodium chloride from 1 to 5 %. The maximum overlapping obtained with 5 % sodium chloride is 1 unit of pH.

(b) When sodium chloride was added to the variable protein concentration series (table 1) a lowering of the clarification threshold concentration of protein was demonstrated. Thus in the absence of sodium chloride, the clarification threshold occurred with the haemoglobin concentration of 2.5 mg./sq.m. of interfacial area, while in the presence of 1 % sodium chloride, the threshold was found to be 1.7 mg. of protein per sq.m. of interfacial area.

4. *Desorption of protein from the interface*

This has been studied with haemoglobin since observation was facilitated by the pigment. The haemoglobin was adsorbed to a negatively charged interface at a pH of 5.0. The resultant red floccules were washed thoroughly in buffer solution at pH 5.0. They were then transferred into a buffer solution at pH 10.0. Redispersal of the emulsion occurred and the fluid became red. This was centrifuged at 10,000 r.p.m. and the emulsion formed a whitish layer at the top, while the deep red pigment was seen in clear solution below. Adsorption and subsequent desorption readily occurred when concentrated solutions were used. If a threshold concentration of protein was used, desorption only occurred after repeated washing and centrifuging in an alkaline buffer, the emulsion breaking towards the end of the experiment. Since breaking also occurred in control emulsions repeatedly centrifuged without protein at pH 10, it was concluded that the stabilizing soap film was removed by this procedure. The desorbed pigmented material has been studied in order to determine what changes, if any, have occurred in the molecular structure as a result of adsorption to and desorption from the oil/water interface. The results of these investigations are shown in table 3. The sedimentation diagram obtained from

TABLE 3. EFFECT OF REVERSIBLE ADSORPTION ON THE HAEMOGLOBIN MOLECULE*

	before adsorption	after desorption
pigment colour	red	red
spectral bands	two well-marked bands at 578 and 540	three very faint bands, one in red and two in green part of spectrum
effect of reduction:		
colour	purple	reddish brown
spectral bands	one broad band at 565	two marked bands at 558 and 520
effect of addition of 1.5 M-phosphate buffer at pH 6.8	pigment freely soluble	pigment precipitated
effect of heating	no change	red colour changes to brown
effect of subsequent cooling	no change	brown colour changes to red
effect of pH change 8-12	no change	reversible red to brown colour change
†sedimentation constant $S_{90}^{\circ} \times 10^{13}$	4.48	3.9
†molecular shape or hydration f/f_0 (in ultracentrifuge)	1.16	1.33
molecular weight by osmotic pressure	68,000	68,000
conclusion	pigment: oxyhaemoglobin	pigment: parahaematin

* Blank results with the stabilizing agents without the oil were carried out and gave no significant results at the concentrations used.

† See figure 5 on ultracentrifuge results.

the ultracentrifuge is shown in figure 5 (plate 5). The conclusion arrived at from these investigations was that human and sheep haemoglobin was changed by adsorption into pure soluble parahaematin (Keilin 1926).

5. *Preferential adsorption of protein from a mixture of two proteins*

A mixture of haemoglobin and albumin was taken and placed in a buffer solution at pH 5.5. Flocculation occurred and the floccules were red in colour; the subnatant fluid was decolorized. When these floccules were removed by centrifugation and further emulsion was added to the subnatant fluid no fresh flocculation occurred, but if the pH of the subnatant fluid was changed to 4.6, white flocculation readily occurred owing to adsorption of the albumin. The proteins can be readily desorbed from the oil/water interface by transference into an alkaline buffer.

DISCUSSION

It has been shown that the results of a study of molecular interactions at an air/water interface can be directly applied to phenomena occurring at an oil/water interface, or at an emulsion particle interface (Schulman & Cockbain 1940). Nearly all proteins spread, or can be made to spread, by suitable dispersion in solutions of alcohols and salts, at the air/water interface and better at the oil/water interface in the form of a monolayer over a large range of pH in the underlying solution. This process entails a radical alteration in the structure of the protein molecule, which has been considered by many workers in this field to be an irreversible change (Langmuir & Waugh (1940); Langmuir & Schaefer (1939)). The experimental results show that adsorbed proteins can readily be made to go back into solution again, although the structure of the adsorbed protein in the various cases quoted has still to be established. The protein molecule has three main configurations. The structure of the regenerated protein in relation to the other two forms, and the biological activity of a protein molecule in these three configurations are of considerable interest.

Form I. Before adsorption

There is evidence from ultracentrifuge studies that the protein molecule in solution is globular or nearly globular in form, with the ionized carboxyl and amine groups at the end of short hydrocarbon chains orientated into the aqueous solution, and the associating non-polar side chains orientated towards the centre of the large molecule. The charge and isoelectric points of the protein molecule are directly related to the ratio of the positive NH_3 and negative COO ionic groups at the end of their side chains.

Thus on the acid side of this isoelectric point, suppression of the ionization of the carboxyl ion takes place and the protein molecule behaves as a positively charged colloid. Conversely on the alkaline side the protein behaves as a negatively charged colloid.

Form II. Adsorbed protein

When a protein molecule of the above description comes into contact with an oil/water or air/water interface, the non-polar portions of the molecule will be separated from the polar and ionic groups owing to the strong asymmetric field at the interface. Thus, the molecule with other unrolled molecules, forms a triplex monolayer. The non-polar side chains will be pulled towards the oil phase or air, and the side chains with the ionic or polar groups towards the water phase, leaving the keto-amido backbone in between. The protein molecules will adlineate to one another according to the stresses placed on the valence linkages by this separation of the non-polar portions (due to the asymmetric field at the interface) and the intramolecular association of these groups (Schulman & Cockbain 1939).

The approach of a protein molecule to a charged oil/water interface is more specific than at the uncharged oil/water interface. At the air/water interface where the surface is covered by a monolayer which does not interact with an injected protein solution, no change as measured by surface potentials, surface pressure, surface viscosity, or rigidity takes place. Examples are: serum proteins injected under lecithin (Hughes 1935) or long-chain ester films (Schulman, unpublished). If the proteins can associate with the film-forming molecules, penetration of the interfacial films takes place with radical changes in the above-mentioned film properties (Schulman, unpublished).

It is a known fact that if protein is adsorbed from a solution to the interface of either oil/water or air/water, the adsorbed film can become thicker than that of a monolayer obtained by the surface-spreading technique. The structure of the proteins in these thick layers is of great interest. Whether these thick layers are unrolled proteins in a laminated monolayer form, or due to adsorption in a globular form, or primarily a monolayer followed by a globular adsorption, will be discussed later (Danielli & Davson 1943).

Form III. Desorbed protein

If surface pressure is exerted on an interfacial protein film, besides increase in surface concentration, two things may occur; the molecule may be forced back into solution again, or it may become so involved with its neighbouring molecules that the interfacial film collapses, forming pleats and rolls. These pleats and rolls, having a hydrophilic core with the non-polar side chains on the outside, can be spun off the surface as threads which do not spread again on the aqueous solution (Schulman & Cockbain 1939). The ease with which a protein molecule can be pushed back into solution again is dependent on its size and on the lipid content of a mixed film. Thus it has been shown that, in a 20 % cholesterol mixed protein film, the protein molecule can be readily forced back into solution again without the film crumpling (Schulman & Rideal 1937). Desorption of a protein molecule from a charged oil/water interface can be brought about in a much more convenient way than at the air/water or uncharged oil/water interface. This is achieved by simply suppressing the ionization of the associating group of opposite sign and the bringing into action the repulsive

forces between ions of like sign. Another method is to break the emulsion by removal of, or interference with, the stabilizing agent. The structure of the desorbed protein has been investigated by determination of its molecular weight, shape (sedimentation constant), and, if pigmented, by its absorption spectrum and, if biologically active, by comparable biological tests with the unadsorbed protein.

It will be seen from the experimental results that these general principles of the behaviour of protein molecules at an oil/water interface can be applied to the study of the reactions which occur when protein solutions are mixed with finely dispersed oil-in-water emulsions. If an emulsion is used in which the oil particles are stabilized with a soap so that the oil globule carries either a positive or a negative charge, adsorption of the protein to the interface is indicated by flocculation of the emulsion particles. This flocculation is due to the reduction of the surface charge on the oil particle. Brownian movement ceases and repellent forces between individual particles no longer exist. The oil particles will, however, remain as separate and discrete globules provided that there is a sufficiently large concentration of protein present.

A number of factors affecting the adsorption of protein, consequent flocculation and the stability of the flocculated particles have been studied from the experimental results. The most important factors can be seen to be: The reaction of the continuous phase, the relative concentrations of protein to interfacial area, the nature of the stabilizing agent and the presence of electrolytes, and the molecular weight of the protein.

Effect of reaction of continuous phase

If protein solutions are taken at concentrations above 2 mg./sq.m. of emulsion interface over the whole range of pH from 4 to 10 for negatively and positively charged emulsions, adsorption, as signified by flocculation of the emulsions, occurs at pH levels below the isoelectric point for negative interfaces, and above this point for positive interfaces. No flocculation occurs above the isoelectric point of the proteins for negative interfaces or below for positive interfaces.

For serum proteins with a molecular weight up to approximately 70,000, the change-over from adsorption to no adsorption over a pH range is very sharp. With high molecular weight proteins such as γ globulin there is an overlap. This might be due to the size of molecule giving it extra Van der Waals associating forces as compared with a smaller molecule; this would enable it to adsorb against a small surface charge. The overlap with γ globulin is about 1.5 pH units on either side of its isoelectric point of pH 6.8. It will be seen that serum albumin in distilled water gives a sharp change-over at the isoelectric point of pH 4.65. The addition of electrolyte to the albumin creates a flocculation overlap at the isoelectric point similar to that seen with γ globulin 0.6 % of NaCl solution. A pH shift of 1 on either side of the isoelectric point is obtained with a 5 % concentration of sodium chloride. This may be due to the effect of electrolyte reducing the ζ potential and thus extending the flocculation range.

The effect of protein concentration

The initial flocculation of the emulsion occurs at a protein concentration of approximately 1 mg./sq.m. of oil/water interface. This concentration is that at which a protein film forms a coherent monolayer at the air/water interface. Alexander & Teorell (1939) show from surface studies of proteins at the uncharged oil/water interface that the oil can expand the protein monolayer uniformly beyond the point of a coherent monolayer which would be formed at an air/water interface. These expanded protein films are fluid but the gelation point remains the same at air/water and oil/water interfaces. The flocculated emulsion with this low protein concentration breaks up into large oil droplets (figure 4, plate 5), and it is significant that the protein monolayers at low surface pressures are fluid and possess low viscosity as compared with a protein film at its equilibrium spreading pressure where it is a strong gel. Thus the oil droplets have become discharged by adsorption of the protein monolayer, but the protein coat around the droplets at this surface concentration is too fragile to stop the discharged oil droplets from coalescing. At a protein concentration of 2.5 mg./sq.m. of interface a firmly gelled monolayer is formed. The emulsion is flocculated in large clumps, each droplet remaining discrete, no free oil being visible and the supernatant solution becoming completely clear of the emulsion droplets which float to the top of the solution as a white layer. If emulsion is mixed with a protein concentration series, this clear tube with stable flocculated emulsion droplets rapidly becomes apparent. This concentration is approximately 0.15 % of protein per 1 c.c. of 5 % emulsion with 0.6 sq.m. of interface. This gives a figure of 2.5 mg. of protein/sq.m. which agrees well with that obtained for a strongly gelled protein monolayer compressed to its equilibrium spreading or collapse pressure (Schulman & Cockbain 1939).

Further, an adsorption of protein in globular form can here be ruled out since the smallest globular unit would have a diameter of 55 Å which is five and a half times greater in thickness than that obtained for the monolayer at 1 mg./sq.m. and some two or three times thicker than the strong gelled monolayers giving maximum clarification of the emulsion. The emulsion particles have a mean diameter of 0.5μ and one rarely obtains any particles in the emulsions used above 1μ .

It can be shown that an emulsion will go on adsorbing protein beyond the monolayer concentration in a manner similar to the adsorption of the strong and thick gel films obtained from a floating protein solution at the air/water and oil/water interface. This can be demonstrated more readily with pigmented proteins. Thus, the concentration of haemoglobin which can be removed by an emulsion to give a colourless solution is of the order of 10 monolayers in thickness. This can only be obtained by the use of a strong protein solution, i.e. 2 %. The structure of these films of adsorbed protein is at present not quite clear. Whether they are laminated monolayers comparable to soap films or adsorbed globular proteins must be established by work on biologically active proteins. The first view would appear more in line with the available facts since adsorbed haemoglobin comes off as para-

haematin, showing that a change has occurred in the protein molecule which would be unlikely if it had been adsorbed in its globular form. This denaturation may, however, only occur with proteins which possess large oil soluble pigmented prosthetic groups. These might irreversibly distort the protein portion of the molecule by their separation at the oil/water interface.

Desorption of protein from charged oil/water interface

The experimental results described show that a protein molecule can only approach a charged oil/water interface when the charges of the two colloids are of the opposite sign, and that the protein molecule occupies the same area at the interface as a protein molecule in the form of a monolayer at the uncharged oil/water or air/water interface. If a protein such as haemoglobin is adsorbed at pH 5.0 on a negative emulsion or at pH 8.5 on a positive interface, and the pH of the solution is now changed so that the sign of the two colloids is the same, repulsion takes place between the adsorbed protein layer and the adsorbed emulsion stabilizing agent (i.e. at pH 10 and 4.6 respectively). With the pigmented haemoglobin the desorption of the protein back into solution again can be readily observed. The flocculated emulsion, which was red in the case of haemoglobin, becomes white again and redisperses to form a fine emulsion. The solution which was originally red becomes decolorized when the emulsion is added under adsorption conditions and the emulsion becomes red. This agglutinated emulsion can be washed with the appropriate buffer solution and then centrifuged. When it is placed in a buffer solution of the opposite reaction, repulsion occurs between the ions of the stabilizing soap film and the ionized groups of the protein molecules in the mixed lipoprotein monolayer at the emulsion interface. Under these conditions the protein molecule redisperses into solution again, and, in the case of the pigmented proteins, colours the solution. That this interfacial film is a mixed monolayer can be seen from penetration experiments at the air/water interface by injecting protein solutions under monolayers of ionizing compounds (Schulman, unpublished). Work with the ultracentrifuge (figure 5, plate 5) suggests that the molecular weight and shape of the desorbed protein, parahaematin, represents a modified haemoglobin in which the position of the pigmented prosthetic groups and its attached groups have changed. This might be due to the reorientation of the porphyrin group on adsorption at the oil/water interface, so that on desorption it associates non-specifically with the protein molecule. From the osmotic pressures and membrane potentials as measured by G. S. Adair, the evidence shows that the charge but not the molecular weight has changed from that of the original haemoglobin.

Preferential adsorption of proteins from mixed solutions

From a study of figure 1 it can be seen that at pH 5.5 on a negative emulsion, haemoglobin will be adsorbed, but serum albumin will not. In a mixed protein solution at pH 5.5, haemoglobin should therefore be adsorbed leaving albumin behind in the solution. This occurs readily; the emulsion at this pH agglutinates into

red floccules. When this flocculated red emulsion is centrifuged off and fresh emulsion is added, it remains suspended. But this emulsion gives a white flocculation if the pH of the solution is now moved to pH 4.6 showing that the albumin has now become adsorbed.

These adsorbed proteins on the two emulsion fractions can now be desorbed by placing them in an alkaline solution of pH 10. Thus it can be established that proteins can be preferentially adsorbed out of mixed solutions as a function of their isoelectric points, the sign of the charge on the oil/water interface, and the pH of the solution, and can readily be desorbed. Adsorption and desorption with large foreign prosthetic groups results in some change in molecular structure (figure 5, plate 5 and table 3), but it is possible that other protein molecules may be adsorbed and desorbed without any change in structure.

Work on possible changes in biological activity of ferments adsorbed in the form of monolayers and multilayers at solid interfaces has been carried out by Langmuir and his co-workers. This work has been discussed and enlarged upon by Lawrence, Miall, Needham & Shih-Chang Shen (1944).

The standardized flocculation technique can be used for the study of certain properties of proteins or for the separation of some protein mixtures. It is also being used for the measurement of interfacial area in emulsions and for the determination of film structure at the oil/water interface. This method is being employed in the study of fat particles during and after absorption from the intestine, and in the investigation of the factors concerned in the stability of particulate fat in the blood. The whole of this work is being continued and extended in order to determine the possible changes in the relationships between proteins and interfaces *in vitro* and *in vivo*. Such changes undoubtedly occur in immunological and other reactions, and it is hoped that the study of the simple system described in this paper may be of assistance in the elucidation of more complex biological phenomena of a similar type.

We are indebted to Professor H. R. Dean for much help and encouragement in the initial experiments. We should like to thank Dr G. S. Adair, F.R.S., for assistance with the experiments on haemoglobin, Dr P. Johnson who carried out the centrifuge experiments on the haemoglobin fractions, W. J. Pardoe for assistance in the preparation of figure 1 and photographs, and Professor E. K. Rideal, F.R.S., for much helpful criticism and advice. We are also indebted to the Sir Halley Stewart Trust for financial assistance in carrying out this work. Thanks are due to Mrs G. S. Adair for giving us fractions of the serum proteins.

REFERENCES

- Adair, G. S. & Robinson, J. 1930 *Biochem. J.* **24**, 993.
 Alexander, A. E. & Teorell, T. 1939 *Trans. Faraday Soc.* **35**, 727.
 Clark 1928 *The determination of the hydrogen ions*. Baillière, Tindal and Cox.
 Danielli, J. F. & Davson, H. 1943 *Permeability of natural membranes*. Cambridge.

- Elkes, J. J., Frazer, A. C. & Stewart, H. C. 1939 *J. Physiol.* **95**, 68.
Frazer, A. C. 1943 *J. Physiol.* **102**, 306.
Frazer, A. C., Schulman, J. H. & Stewart, H. C. 1944 *J. Physiol.* **103**, 306.
Frazer, A. C. & Stewart, H. C. 1940 *Brit. J. Exp. Path.* **21**, 361.
Frazer, A. C. & Walsh, V. G. 1933 *J. Physiol.* **78**, 467.
Frazer, A. C. & Walsh, V. G. 1934 *Brit. Industr. J.* **424**.
Frazer, A. C. & Walsh, V. G. 1939 *J. Pharmacol.* **67**, 476.
Hughes, A. H. 1935 *Biochem. J.* **29**, 430, 437.
Keilin, D. 1926 *Proc. Roy. Soc. B*, **100**, 187.
Langmuir, I. & Schaefer, V. J. 1939 *Chem. Rev.* **24**, 181.
Langmuir, I. & Waugh, D. F. 1940 *J. Amer. Chem. Soc.* **62**, 2771.
Lawrence, A. S. C., Miall, M., Needham, J. & Shen, S.-C. 1944 *J. Gen. Physiol.* **27**, 233.
Schulman, J. H. & Cockbain, E. G. 1939 *Trans. Faraday Soc.* **35**, 1266.
Schulman, J. H. & Cockbain, E. G. 1940 *Trans. Faraday Soc.* **36**, 651.
Schulman, J. H. & Rideal, E. K. 1937 *Proc. Roy. Soc. B*, **122**, 29.
Svedberg, T. & Pedersen, K. O. 1940 *The Ultracentrifuge*. Oxford Univ. Press.

Supersonic dispersion in gases

II. Air containing water vapour

BY M. MOKHTAR AND E. G. RICHARDSON

King's College, Newcastle-upon-Tyne

(Communicated by W. E. Curtis, F.R.S.—Received 4 December 1939)

INTRODUCTION

It has long been known from incidental observation that humidity affects the propagation of sound in air. Not only is there an increase in the velocity of sound waves, but, refraction due to irregularities apart, there is a change in the attenuation, so that distant sounds may often be heard more clearly under conditions of high humidity than when the air is dry. Similar effects of humidity on the absorption of sound in reverberation chambers were reported by Sabine (1929), but it was left to Knudsen (1931, 1933) to show that the absorption of sound in air as derived from reverberation experiments at audible frequencies reaches a maximum (as the humidity increases) at first and then falls, and that when the frequency is changed the position of the maximum also moves. This anomalous behaviour of sound in moist air was pursued into the supersonic range of frequencies by Pielemeier. In 1937 (after the present work was commenced) he reported two humidity regions exhibiting abnormal velocity, with sources of sound covering the range 500–100 kc./sec. This year (1939) he has compared his measured velocities with those given by a formula of Miller, based on outdoor experiments, which indicates a steady rise of velocity with humidity. Whereas the earlier experiments of Reid (1930) and Ishii (1935) had shown no measurable departure from the Miller formula, Pielemeier attained values definitely higher at certain values of humidity and temperature.

Measurements of absorption in moist air in the supersonic region are scanty, and the methods employed open to criticism (*vide infra*). Rogers (1934) reported maxima of absorption at 45% humidity in air (mixed also with carbon dioxide) at one frequency (410 kc./sec.), while Pielemeier showed the existence of a critical vapour pressure (for maximum absorption) which was a function of the frequency, similar to that deduced by Knudsen for lower frequencies.

In view of the uncertainty and conflict between the existing results it was decided in 1937 to adopt the hot-wire method, previously used by one of us to measure dispersion in pure gases (Richardson 1934), to measure the wave-length and attenuation coefficients of the radiation from sources covering a range of supersonic frequencies in air containing water vapour. It may be mentioned that dispersion has been detected in other gaseous mixtures, but since air and water vapour formed a typical and common mixture, it was decided to confine the experiments to these gases. Though dry air is, of course, itself a mixture, careful experiments have shown no dispersion over the range of frequencies we intended to employ (cf. Parker 1937).

APPARATUS AND METHOD OF MEASUREMENT

The general form of the apparatus was the same as that used in the first research (Richardson 1934, figure 2). A quartz oscillator produced standing waves between its radiating face and a movable ebonite reflector, while a platinum hot-wire detector could be passed along the wave system so set up, and the amplitude measured from point to point. One improvement which was made within the vessel was to have the hot wire so mounted that it could be pushed along rails in one direction by the head of a screw of a dividing engine, instead of being mounted on a threaded rod directly, as shown in the sketch of the original apparatus. This change was made because it appeared that aberrations in the readings periodic with the thread of a micrometer screw could be introduced by the first arrangement, and backlash was sometimes evidenced by slight disagreements between the direct and return readings of the hot wire. Apart from this and the introduction of a thermometer, the gas container and its appurtenances, the method of taking hot wire or reflector readings and calculating therefrom the propagation constants for the gas remained as described in the first paper.

The methods used by other workers, though excellent for wave-length determinations, suffer from the disadvantage when used to obtain the attenuation constants in gases that they involve the reaction on a nearby quartz oscillator of the same frequency, or on the sender quartz when a reflector is moved, or again, on a torsion or reactive vane placed in the path of the supersonic beam. Movement of the body receiving the reaction is bound to upset, to some extent, the supersonic source and its radiated beam. It is for this reason that the present authors favour an apparatus in which the standing wave system is kept unchanged, while the amplitude

from place to place is measured by some detecting device presenting a very small surface to the beam. In this case it was a length of platinum wire one-thousandth of an inch in diameter. There have been occasions, however, with the crystals of the highest frequency, when the amplitude

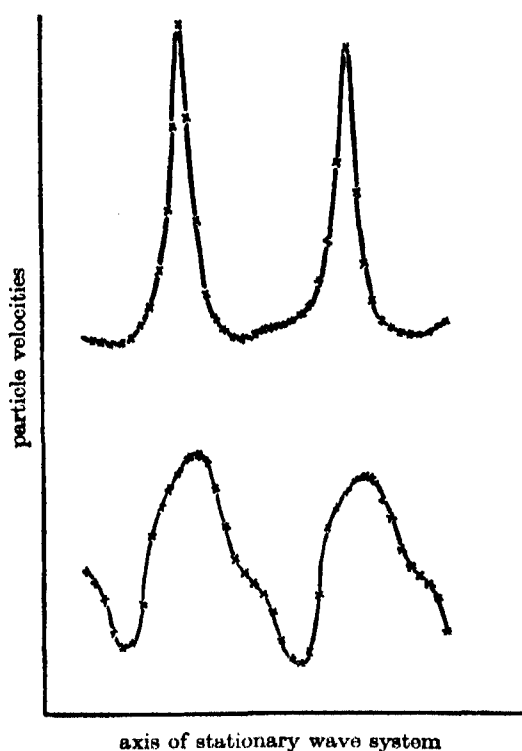


FIGURE 1. Standing waves in a supersonic beam in moist air at 96.8 kc./sec.; above, stationary hot wire and moving reflector; below, stationary reflector and moving hot wire.

in the supersonic waves was too weak to be detected by this means. In such cases the hot wire was kept still at a suitable distance from the quartz and the reflector moved. In some instances, for the sake of comparison, we employed the two methods consecutively under identical conditions. Typical plots of supersonic amplitude, as recorded by the hot wire in the two methods, are shown in figure 1. (It must be remembered that the hot wire does not discriminate phase, and to compare these with stationary waves, as usually pictured, one must imagine alternate segments replaced

by their images in the axis of zero amplitude.) The moving hot wire and stationary reflector plot gives a good approximation to simple harmonic motion (apart from the effects of attenuation, and a suspicion of the octave of the fundamental frequency), whereas with a stationary hot wire and moving reflector sharp-pointed peaks are obtained, which, if used to calculate absorption, usually gave a higher value. Whenever possible then, the reflector was kept still for absorption measurements, though the moving reflector was preferred for velocity measurements, owing to the sharper peaks.

A modification of the hot-wire circuit was introduced, by which the wire with its heating battery was made to form the grid bias of a valve so chosen that the anode current of the valve was directly proportional to the velocity of the draught on the hot wire and hence to the supersonic velocity amplitude. Over the range required in these experiments a variable-mu valve (VS24) was found to fulfil this requirement. The relative amplitudes at points in the standing wave system could then be read off directly from a galvanometer in the plate circuit of this valve.

The humid air to be experimented on was prepared in the desired state, passed into the acoustic interferometer, and its humidity measured on leaving. The air was first freed from carbon dioxide by passing it through potassium hydroxide and subsequently bubbling it through concentrated sulphuric acid. This was done because small concentrations of carbon dioxide are known to effect some dispersion, and it was not desirable to complicate the experiments by double dispersion (cf. Rogers 1934). From the sulphuric acid bottle the gas could pass by alternative routes. One gave it further dryness by passing it over phosphorus pentoxide. The other led through aqueous solutions of salts whose saturation vapour pressure gave the air the approximate humidity desired.

The whole of this apparatus, as well as the interferometer, was in enclosures which could be raised to any temperature up to 80°C by means of external heating coils, and such portions of the connecting tubes as could not be accommodated in the chamber were heavily lagged with asbestos heated by coils to somewhat higher temperatures than the air passing through, to avoid risk of condensation.

After leaving the preparation chamber, the gas could pass directly through the hygrometer, for calibration purposes, or—when the supersonic measurements were being made—first through the interferometer and then through the hygrometer. This procedure was adopted as it was inconvenient to evacuate the interferometer before introducing the gas. The known vapour pressures over the solutions could, however, be used to calibrate

the hygrometer since, when the interferometer was out of circuit, the remaining connexions were completely airtight.

The hygrometer was also of the hot-wire type, in which the electrical resistance of a thin platinum wire fitted axially to a tube is used to measure

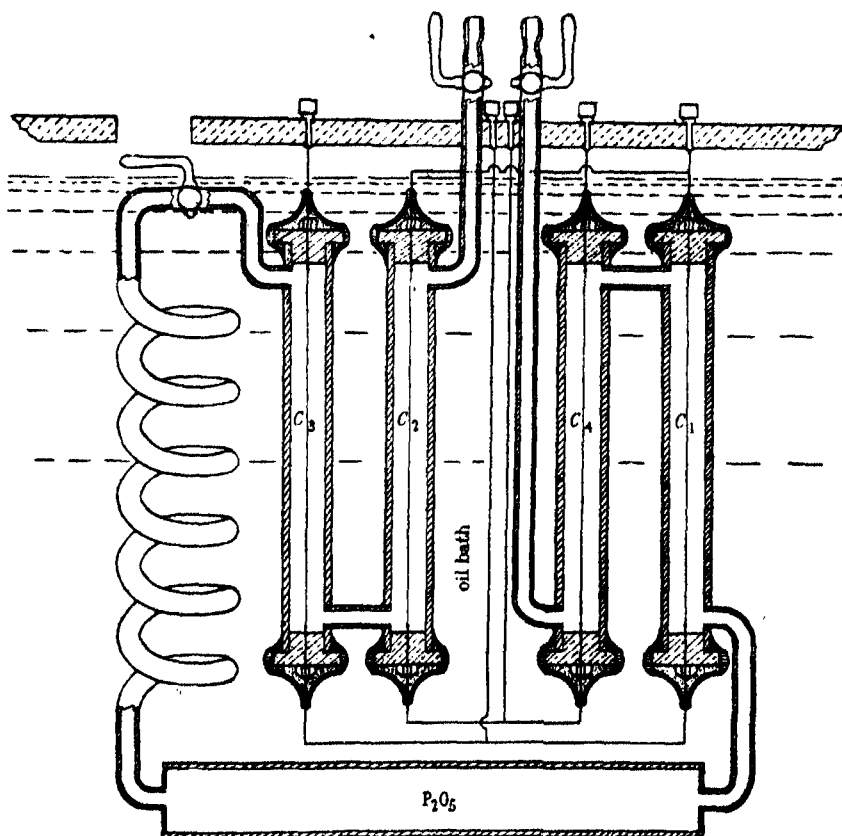


FIGURE 2. Hot wire hygrometer.

the thermal conductivity of any gas in the tube. As the conductivity of air changes but slowly with the addition of water vapour, the single tube type of katharometer did not give sufficient sensitivity. Instead, four conductivity cells with associated hot wires were fitted up, two to contain moist and two dry air. These four wires were connected in a Wheatstone bridge so that each similar pair formed two conjugate arms of the bridge, thus doubling the sensitivity. The final design of the hygrometer is shown in figure 2. Each cell consists of a platinum wire, 0.001 in. in diameter,

mounted in a copper tube 1 cm. diameter and 10 cm. long. At its ends the wire is soldered to two thin brass rods, each brought through an ebonite cap and held under tension by nuts, sealed in position with sealing wax and covered with a cement (Insalute). The whole is mounted in a thermostat kept at 60° C.

The humid air passes through two of the cells in succession, then through the phosphorus pentoxide tube, and through the remaining pair of cells. The object of using the humid air after drying it as a standard gas is to make the readings independent of impurities other than water vapour, and of casual changes of pressure and temperature in the hygrometer. It is, of course, very necessary in using such an apparatus to ensure that all four cells remain at the same temperature. The calibration curve was obtained in the form of the positions of the galvanometer lead on a slide wire placed between two of the bridge arms, for a series of known vapour pressures over the aqueous solution employed.

RESULTS

Measurements of amplitude were obtained by the hot wire in the standing waves produced by supersonic sources at five different frequencies. To obtain the propagation constants in the gas it is sufficient, as shown in the first paper, to measure the position and amplitude at successive maxima and minima in the pseudo-stationary waves set up between the source and the reflector. By plotting the logarithm of the amplitude against distance from the source, the (amplitude) absorption coefficient μ is obtained. From the average separation of peaks the half wave-length $\frac{1}{2}\lambda$ is obtained. Knowing the frequency from the readings of a wave-meter (guaranteed by the makers to be correct to 0.1 %), the velocity is at once given. Typical results are shown on figure 3 of the drift of the maxima and minima in the standing waves, and the log amplitude:distance curve whose slope determines μ .*

At higher frequencies a larger number of peaks were covered within the usual separation of sender and crystal (about 10 cm.). A cyclic variation in the amplitude which was found to have a "wave-length" equal to the pitch of the reflector and hot-wire screws was eliminated when these were pushed continually in the one direction by the dividing engine before mentioned.

* Further details of the results and methods of calculation are given in an Appendix, which has been deposited with the Society.

The complete set of absorption coefficient:humidity curves for the five frequencies is shown on figure 4, while figure 5 shows the corresponding velocity:humidity relationship. (On classical theory the latter curves would run steadily without points of inflexion from a value of 331.5 m./sec. for dry air to 337 m./sec. at a vapour pressure of 10 cm. of mercury.) The

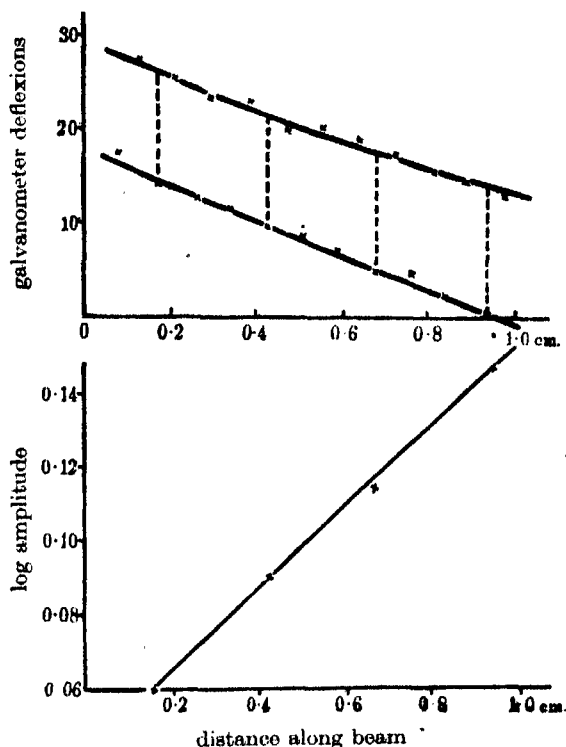


FIGURE 3. Maxima and minima and attenuation of amplitude in standing wave system (200 kc./sec. in air at 20° C and 1.76 cm. water vapour pressure).

humidity at which the absorption is greatest corresponds closely to that at which the velocity shows its greatest departure from the theoretical value. The results are summarized in table 1.

The experimental values of μ (shown in brackets) at the lowest frequency have been corrected in the light of the data obtained in cross-traverses of the hot wire through the supersonic field (vide next section), which indicate the proportion of energy diffracted out of the central supersonic beam. At higher frequencies the correction falls within the experimental error in the determination of μ and so has been neglected.

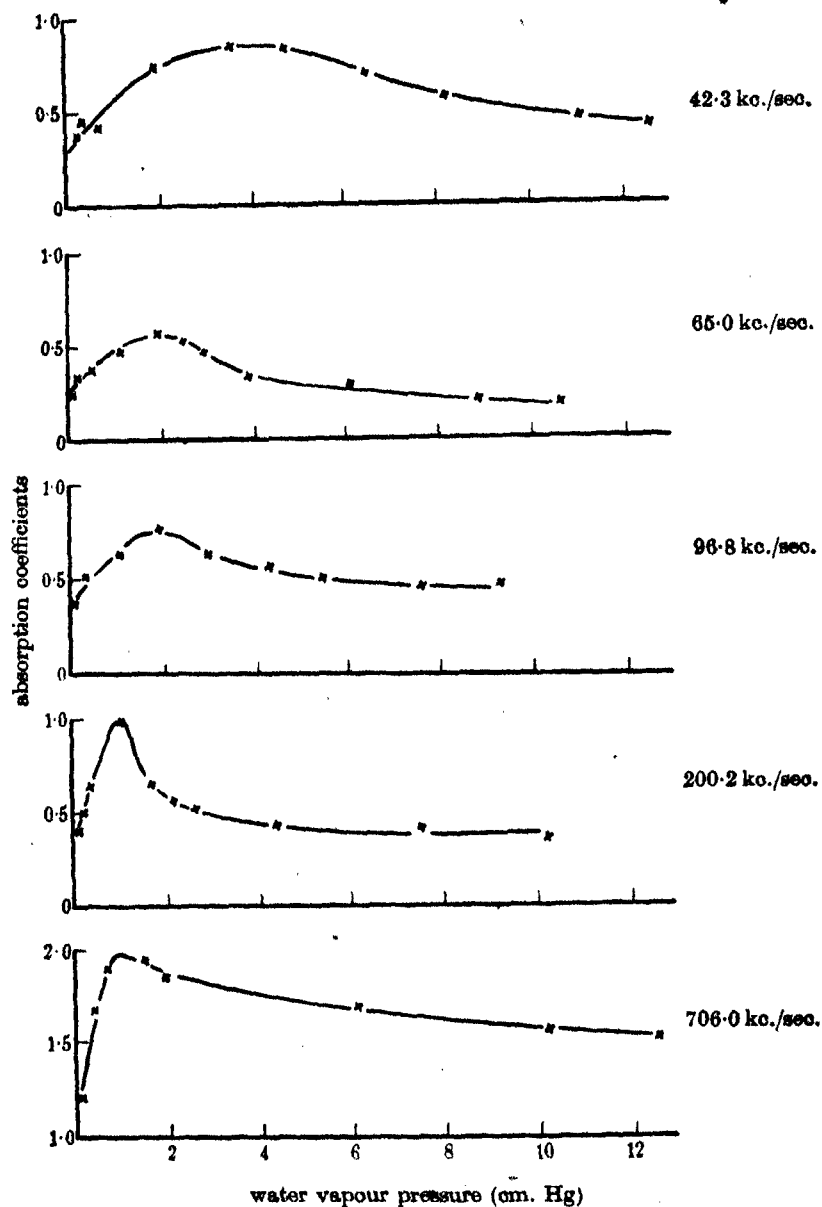


FIGURE 4. Supersonic absorption coefficients at various frequencies.

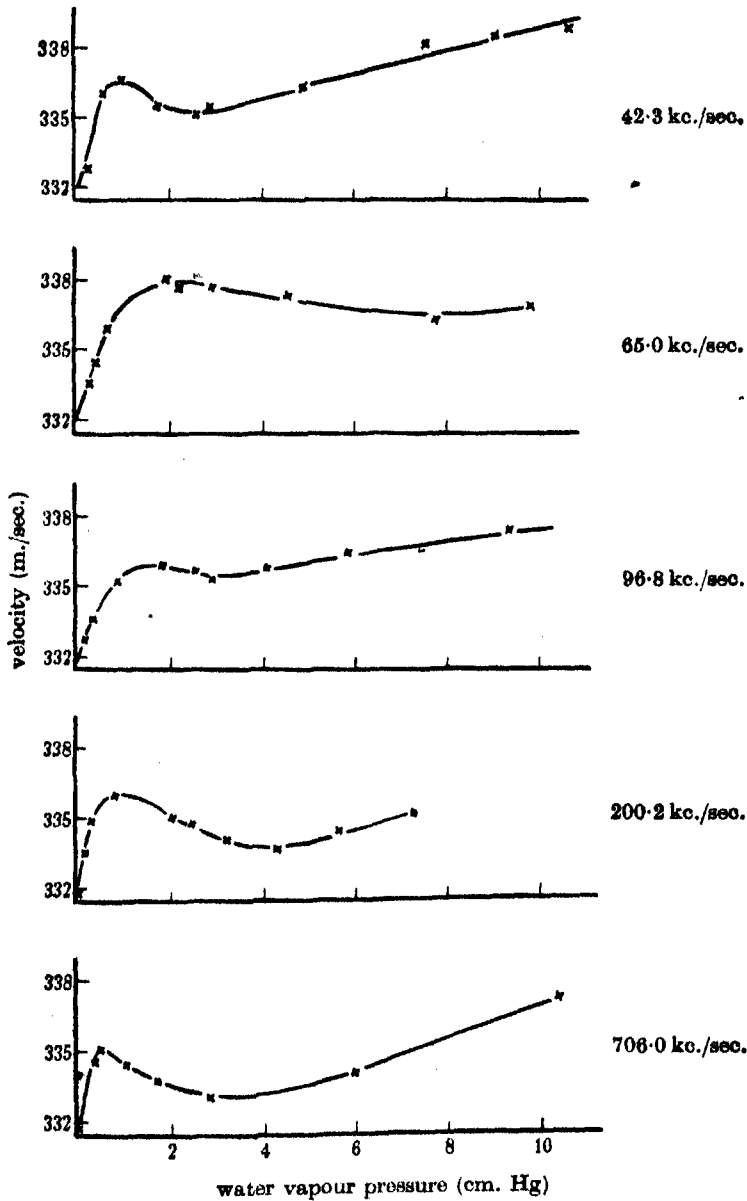


FIGURE 5. Supersonic velocities (reduced to 0° C) at various frequencies.

TABLE 1

Frequency kc./sec.	μ		$\mu\lambda^{**}$		V m./sec.	Vapour pressure (cm. Hg)	
	Dry air	At max.	Dry air	At max.		At max. μ	At max. V
42.3	0.15	0.7	0.09	0.45	337.5	3	1
	(0.3)	(0.85)					
65.0	0.25	0.55	0.065	0.15	337.5	2	2
96.8	0.35	0.75	0.04	0.09	336	1.7	1.7
200.2	0.45	1.0	0.01	0.025	336	1	0.8
706.0	1.2	1.9	0.003	0.005	335	0.8	0.6

* The Stokes-Kirchhoff formula would give $\mu\lambda^2 = 0.0004$.

The following general conclusions may be drawn:

- (1) The velocity in dry air is independent of frequency (cf. Parker 1937).
- (2) The measured absorption coefficients in dry air are several times larger than those calculated from the Stokes-Kirchhoff formula, and while this would have μ depend on the square of the frequency, the results indicate an approximately linear dependence. Consequently $\mu\lambda^2$ decreases as the frequency increases, instead of remaining constant.
- (3) In humid air μ reaches a maximum, two or three times its value in dry air, at a vapour pressure which decreases as the frequency increases. (The absorption curves of figure 4 can be nearly superposed if plotted against the logarithm of the vapour pressure divided by the frequency.)
- (4) The maximum of dispersion in the velocity decreases as the frequency increases.

DISCUSSION OF RESULTS

Before discussing the import of these results, some investigation is needed of the disposition of the supersonic field in front of the quartz. We have, in fact, assumed in the foregoing calculations that the radiation all falls on the reflector and is there sent back towards the source in the form of a sensibly plane wave. The criterion for the production of such a beam without marked diffraction is, as in the corresponding optical case, the relative size of wave-length and radiating face of the sender. For the high frequencies, where one had wave-lengths of the order of fractions of a millimetre coming from a quartz disk about 2 cm. wide, the amount of energy diffracted out of the central beam was very small, but at the lowest frequency (provided by a quartz slab with radiating face 2 cm. wide, delivering radiation with a wave-length in air of 0.85 mm.) diffraction was

evident first by reflexion of the radiation from the walls of the glass tube—reduced by covering it internally with gauze—and an exaggerated absorption coefficient in dry air due to departure by the supersonic radiation from plane waves. In spite of this diffraction at the lowest frequencies, the increased absorption shown in humid air at these frequencies may still be regarded as a specific property of the humidity and not of the apparatus. It is possible, however, that the abnormal absorption with certain concentrations of water vapour may be due to scattering of the radiation, as was previously found in like circumstances when the interferometer contained pure carbon dioxide (Richardson 1934). To test this, and to verify that the supersonic field in dry air was the same as that which one would expect on theoretical grounds, the field in front of certain of the sources when radiating through a linear slit was explored by means of the hot-wire detector. For this purpose the reflector was removed, the interior of the walls and the far end of the tube covered with felt, and the field in front of the slit orifice traversed across plane sections of the gas perpendicular to the supersonic rays and with the hot wire parallel to the long edges of the slit. This was done both in dry air and afterwards with a moisture content at the concentration corresponding to maximum absorption for the frequency concerned. It was apparent that while the field in humid air does show greater divergence from theory than that in dry air there is not the marked irregularity in the field which led to the “absorption” in carbon dioxide being ascribed to scattering. We should indeed not expect so much divergence from “piston radiation” in this case, for the extent of the recorded dispersion (both of velocity and amplitude) is much less in humid air at its greatest than in carbon dioxide in the appropriate frequency range. Incidentally, the making of the traverses with the hot wire across the supersonic field enabled one to verify that the beam was in fact proceeding out at right angles to the crystal face and axially to the tube, by noting the relative position of the central maximum at various distances from the source. Needless to say, any faulty setting of this type, which would result in the beam not hitting the reflector squarely, could produce a serious error in the calculated results for absorption coefficient.

The significance of supersonic dispersion from the point of view of molecular characteristics has been discussed both in the first paper and elsewhere (Richardson 1938). The argument lies between the relaxation theory, in which a postulated lag in the transfer of energy between the various degrees of freedom of the molecule produces a peak of absorption hand-in-hand with a rise in velocity to a higher and *maintained* value; and

some form of selective absorption, characterized by a sharp rise in velocity followed by a gradual fall to a normal value (cf. anomalous dispersion in optics). It is not the purpose of this paper to enter into a theoretical discussion of the merits of these rival theories. It suffices to point out that the velocity after it has risen to an abnormal value does in fact fall slowly to normal as the frequency rises. So much may be deduced from figure 5, if a succession of values of velocity corresponding to a single value of the humidity is read off from the curves. The drop in velocity prior to the rise at a frequency corresponding to maximum absorption, which should accompany anomalous dispersion, is less evident, because in the supersonic gamut the maximum occurs at low vapour pressure, but velocities *less* than those given by the classical formula were occasionally recorded at very low values of the humidity.

The authors desire to express their best thanks to Professor W. E. Curtis, F.R.S., for his encouragement during the course of this research and for placing the facilities of the laboratory at their disposal, and to Professor E. N. da C. Andrade, F.R.S., for helpful criticism in the later stages of this work, as a result of which they were able to bring greater precision into the measurements.

REFERENCES

- Kinoshita, M. and Ishii, C. 1932 *Sci. Rep. Inst. Phys. Chem. Res., Tokyo*, **19**, 83.
Ishii, C. 1935 *Sci. Rep. Inst. Phys. Chem. Res., Tokyo*, **26**, 201.
Knudsen, V. O. 1931 *J. Acoust. Soc.* **3**, 126.
— 1933 *J. Acoust. Soc.* **5**, 112.
Parker, R. C. 1937 *Proc. Phys. Soc.* **49**, 35.
Pielemeier, W. H. 1937 *Phys. Rev.* **52**, 244.
— 1939 *J. Acoust. Soc.* **10**, 313.
Reid, C. D. 1930 *Phys. Rev.* **35**, 814.
Richardson, E. G. 1934 *Proc. Roy. Soc. A*, **146**, 56.
— 1938 *Phys. Soc. Progr. Rep.* **4**, 67.
Rogers, H. H. 1934 *Phys. Rev.* **45**, 208.
Sabine, P. E. 1929 *J. Franklin Inst.* **207**, 347.
-

Complex potentials in two-dimensional elasticity

By A. C. STEVENSON

(Communicated by G. B. Jeffery, F.R.S.—Received 25 May 1940)

This paper gives an approach to two-dimensional isotropic elastic theory (plane strain and generalized plane stress) by means of the complex variable resulting in a very marked economy of effort in the investigation of such problems as contrasted with the usual method by means of Airy's stress function and the allied displacement function. This is effected (i) by considering especially the transformation of two-dimensional stress; it emerges that the combinations $\widehat{xx} + \widehat{yy}$, $\widehat{xx} - \widehat{yy} + 2i\widehat{xy}$ are all-important in the treatment in terms of complex variables; (ii) by the introduction of two complex potentials $\Omega(z)$, $\omega(z)$ each a function of a single complex variable in terms of which the displacements and stresses can be very simply expressed.

Transformation of the cartesian combinations $u + iv$, $\widehat{xx} + \widehat{yy}$, $\widehat{xx} - \widehat{yy} + 2i\widehat{xy}$ to the orthogonal curvilinear combinations $u_\xi + iu_\eta$, $\widehat{\xi\xi} + \widehat{\eta\eta}$, $\widehat{\xi\xi} - \widehat{\eta\eta} + 2i\widehat{\xi\eta}$ is simple and speedy.

The nature of the complex potentials is discussed, and the conditions that the solution for the displacements shall be physically admissible, i.e. single-valued or at most of the possible dislocational types, is found to relate the cyclic functions of the complex potentials.

Formulae are found for the force and couple resultants at the origin $z = 0$ equivalent to the stresses round a closed circuit in the elastic material, and these also are found to relate the cyclic functions of the complex potentials.

The body force has been supposed derivable from a particular body-force potential which includes as special cases (i) the usual gravitational body force, (ii) the reversed mass accelerations or so-called 'centrifugal' body forces of steady rotation.

The power of the complex variable method is exhibited by finding the appropriate complex potentials for a very wide variety of problems, and whilst the main object of the present paper has been to extend the well-known usefulness of the complex variable method in non-viscous hydrodynamical theory to two-dimensional elasticity, solutions have been given to a number of new problems and corrections made to certain other previous solutions.

INTRODUCTION

The power and elegance of complex variable methods of dealing with two-dimensional problems of hydrodynamics is now well recognized (e.g. see Milne-Thomson 1938). The gain over earlier methods is comparable with the gain in three-dimensional work by a systematic use of vectors, and, indeed, the fact that the symbol i is the vector operator $[\mathbf{k}, \]$ so that the vector velocity \mathbf{q} is linked with a complex velocity $u + iv$ by the relations

$$\mathbf{q} = u\mathbf{i} + v\mathbf{j} = u\mathbf{i} + v[\mathbf{k}, \mathbf{i}] = \{u + iv\}\mathbf{i},$$

gives one reason for this likeness. The student whose chosen branches of study embrace both hydrodynamics and elasticity cannot fail to notice a complete contrast in the approach to two-dimensional problems involving the same boundaries.

For although it is well known that states of plane strain in two-dimensional isotropic elasticity can be determined in terms of functions of a complex variable (e.g. see Love 1927, p. 204), points of application of force corresponding to poles and logarithmic singularities of these functions, which are conserved in conformal transformations, yet such transformations do not transform a boundary free from traction into a boundary free from traction. This defect of correspondence has been regarded as the main difficulty in the way of advance in the theory of two-dimensional elastic systems (Love 1927, p. 214), and the use of complex functions has been practically restricted to the study of isolated nuclei of strain (but see MacGregor 1935, p. 177).

As a consequence, problems of two-dimensional elasticity are usually approached by means of Airy's stress function $\chi(x, y)$, satisfying the biharmonic equation $\nabla^4\chi = 0$ when body forces are absent, coupled with a plane harmonic displacement function $\psi(x, y)$, related to $\chi(x, y)$ by the equation $\nabla^2\chi = \partial^2\psi/\partial x\partial y$ (e.g. see Coker and Filon 1931, p. 130). Many interesting problems have been solved in this way by the aid of appropriate orthogonal curvilinear co-ordinates, as, for example, with polar co-ordinates (Michell 1899, 1901), elliptic co-ordinates (Inglis 1913 *a, b*), and bipolar co-ordinates (Jeffery 1921). But the path to these solutions is by no means short. We have to consider, in terms of the curvilinear co-ordinates, (i) the transformation of stress components, (ii) the transformation of the biharmonic equation, (iii) solutions for χ involving double boundary conditions, (iv) the solution for ψ . Now all these stages involve much partial differentiation, and the algebra is generally very heavy and tedious. Also, unlike Laplace's equation, the biharmonic equation is not invariant in form for these transformations. Further, although the work is sometimes shortened by omitting to consider the displacements, this is a dangerous practice as it sometimes leads to stresses satisfying boundary conditions but which can only arise from physically impossible displacements.

Recognition of some of these complexities have led one or two writers to consider these problems entirely in terms of plane harmonic functions (e.g. see Love 1927, p. 204; Carothers 1920, p. 110; and particularly Bricas 1932, p. 11, who dispenses altogether with Airy's stress function). But these writers make little or no effective use of the complex variable in this connexion, with the result that their methods still retain very much of the

causes responsible for the long and somewhat heavily algebraic path to the required solutions.

What is required is (i) a means of making a rapid change from one set of orthogonal co-ordinates to another, and (ii) a *modus operandi* which does not entail the solution of fresh differential equations whenever the co-ordinates are changed. We shall show that the stress combinations $\bar{x}\bar{x} + \bar{y}\bar{y}$, $\bar{x}\bar{x} - \bar{y}\bar{y} + 2i\bar{x}\bar{y}$ and $\bar{\xi}\bar{\xi} + \bar{\eta}\bar{\eta}$, $\bar{\xi}\bar{\xi} - \bar{\eta}\bar{\eta} + 2i\bar{\xi}\bar{\eta}$ in curvilinear co-ordinates are very simply related and admirably fulfil our first requirement, and we shall further show that the solution of two-dimensional elastic problems for the stresses and displacements, including cases with body force, depend upon two functions Ω and ω of a single complex variable, and which we term complex potentials. The method arises naturally from a consideration of these problems in terms of co-ordinates z , \bar{z} instead of x , y where $z = x + iy$, $\bar{z} = x - iy$, and since it side-tracks the biharmonic equation, it fulfils our second requirement. Like the method of Bricas, to which it is more nearly related than the Airy stress function method, it dispenses altogether with Airy's stress function and the allied displacement function.

The cyclic constants of the complex potentials are shown to occur in (i) the criteria for the displacements to be single-valued or to be at most dislocational in character, (ii) the expressions for the stress resultants round closed circuits, and this narrows down considerably the search for the appropriate complex potentials in any given problem. By considering a number of by now classical problems, notably those in elliptic and bipolar co-ordinates, generalizing and correcting certain other solved problems, as well as by giving the solution of certain new problems, the writer shows that the method of complex potentials represents a very considerable gain in elegance and algebraic simplicity over previous methods.

1. TWO-DIMENSIONAL ELASTICITY

We shall consider the fundamental problem of two-dimensional statical isotropic elasticity to be that of *plane strain*, since the solution of problems of plane stress and *generalized plane stress* can be made to depend largely upon that of plane strain. In plane strain the restrictions upon the displacements u , v , w are

$$u = u(x, y), \quad v = v(x, y), \quad w = 0. \quad (1.1)$$

If δ is the dilatation, and ωk is the rotation, these are given by

$$\delta = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}, \quad 2\omega = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}. \quad (1.2)$$

From (1.1) and the stress-strain relations we have

$$\widehat{xz} = 0, \quad \widehat{yz} = 0, \quad \widehat{zz} = \lambda\delta \quad (1.3)$$

and
$$\widehat{xx} = \lambda\delta + 2\mu \frac{\partial u}{\partial x}, \quad \widehat{yy} = \lambda\delta + 2\mu \frac{\partial v}{\partial y}, \quad \widehat{xy} = \mu \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right), \quad (1.4)$$

where λ, μ are Lamé's elastic constants, so that

$$\widehat{xx} + \widehat{yy} = 2(\lambda + \mu)\delta, \quad \widehat{xx} - \widehat{yy} = 2\mu \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right). \quad (1.5)$$

In a state of generalized plane stress (see Coker and Filon 1931, p. 126) in a plate-like material bounded by $z = \pm c$, we take $\widehat{zz} = 0$ throughout the material, with \widehat{xz} and \widehat{yz} vanishing over the surfaces $z = \pm c$. The displacements \bar{u}, \bar{v} and stresses $\bar{xx}, \bar{yy}, \bar{xy}$ averaged across the thickness $2c$ of the plate, are then functions of x, y only, and are related in precisely the same manner as the actual displacements and stresses in plane strain, i.e. as in (1.4), provided the Lamé constant λ is replaced by a modified Lamé constant λ' given by

$$\lambda' = 2\mu\lambda/(\lambda + 2\mu), \quad (1.6)$$

so that the problems of plane strain and generalized plane stress are mathematically identical to this extent.

The problem of plane stress is defined by

$$\widehat{xz} = \widehat{yz} = \widehat{zz} = 0. \quad (1.7)$$

Here again the remaining stresses are given in terms of the displacements by equations (1.4) when λ is replaced by λ' . They are not, however, independent of z in general, so that the problem of plane stress is less strictly a two-dimensional one. We shall not consider it further here, but remark only that its solution can be effected by the use of complex potentials in much the same way as the problems of plane strain and generalized plane stress.

2. THE COMPLEX DISPLACEMENT

We use a complex displacement D related to the vector displacement \mathbf{D} by the relations

$$\mathbf{D} = u\mathbf{i} + v\mathbf{j} = u\mathbf{i} + v[\mathbf{k}, \mathbf{i}] = \{u + iv\}\mathbf{i} = D\mathbf{i}, \quad (2.1)$$

and, instead of the variables x, y we use the complex variables

$$z = x + iy, \quad \bar{z} = x - iy. \quad (2.2)$$

There will be no occasion to use z hereafter as the third spatial co-ordinate in any section which uses the customary symbol z for the complex variable, so that no confusion need arise from the double use of z . With these variables we have

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial z} + \frac{\partial}{\partial \bar{z}}, \quad \frac{\partial}{\partial y} = i \left(\frac{\partial}{\partial z} - \frac{\partial}{\partial \bar{z}} \right), \quad (2.3)$$

and
$$2 \frac{\partial}{\partial z} = \frac{\partial}{\partial x} - i \frac{\partial}{\partial y}, \quad 2 \frac{\partial}{\partial \bar{z}} = \frac{\partial}{\partial x} + i \frac{\partial}{\partial y}. \quad (2.4)$$

Hence (1.5) and (2.1), (2.4) give

$$2 \frac{\partial D}{\partial z} = \delta + 2i\pi, \quad 4\mu \frac{\partial D}{\partial \bar{z}} = \bar{x}x - \bar{y}y + 2i\bar{x}y. \quad (2.5)$$

3. THE COMPLEX BODY-STRESS EQUATION

We assume the body force per unit mass to be derivable as the negative gradient of a body-force potential V . In plane strain and plane stress one body-stress equation shows that the resolute of body force perpendicular to the x - y plane must vanish, and we suppose in all cases that

$$V = V(x, y) = U(z, \bar{z}); \quad (3.1)$$

also we find it convenient later to introduce a body-force function $W(z, \bar{z})$, such that

$$W(z, \bar{z}) = \int U(z, \bar{z}) dz. \quad (3.2)$$

In particular, we consider two cases of body force included in the form

$$U(z, \bar{z}) = cz + \bar{c}\bar{z} + d\bar{z}\bar{z}, \quad (3.3)$$

where d is real. This makes

$$W(z, \bar{z}) = \frac{1}{2}cz^2 + \bar{c}z\bar{z} + \frac{1}{2}dz^2\bar{z}, \quad \frac{\partial W}{\partial \bar{z}} = \bar{c}z + \frac{1}{2}dz^2. \quad (3.4)$$

If
$$c = -\frac{1}{2}ge^{i\alpha}, \quad d = 0, \quad (3.5)$$

we have the ordinary constant gravitational body force g acting at an angle α with the x -axis. If

$$c = \frac{1}{2}n^2\bar{z}_0, \quad d = -\frac{1}{2}n^2, \quad (3.6)$$

we have the body-force potential of the reversed mass accelerations (treated as body forces) in a steady rotation of the elastic material in its plane about the point $z = z_0$ with constant angular velocity n .

The two remaining body-stress equations in plane strain and plane stress are

$$\frac{\partial}{\partial x}(\bar{x}\bar{x} - \rho V) + \frac{\partial}{\partial y}\bar{x}\bar{y} = 0, \quad \frac{\partial}{\partial x}\bar{x}\bar{y} + \frac{\partial}{\partial y}(\bar{y}\bar{y} - \rho V) = 0, \quad (3.7)$$

and we merely replace the actual stresses by the mean stresses in the problem of generalized plane stress. Multiply the second equation of (3.7) by i and add to the first, and make use of (2.3); we have

$$\frac{\partial}{\partial z}(\bar{x}\bar{x} - \bar{y}\bar{y} + 2i\bar{x}\bar{y}) + \frac{\partial}{\partial \bar{z}}(\bar{x}\bar{x} + \bar{y}\bar{y} - 2\rho U) = 0, \quad (3.8)$$

as the complex form of the body-stress equations in two-dimensional elastic problems. This is entirely independent of the third spatial co-ordinate in the two cases of plane strain and generalized plane stress. Suppose $\bar{p}q_1, \bar{p}q_2$ ($p, q = x, y, z$) are two sets of stresses satisfying (3.8), then the stresses $\bar{p}q_0 = \bar{p}q_1 - \bar{p}q_2$ clearly satisfy the body-stress equations with no body force. Hence if we form a second solution $\bar{p}q_2$ of the equations by superposing a complementary solution $\bar{p}q_0$ upon a particular solution $\bar{p}q_1$, then

$$\bar{p}q_2 = \bar{p}q_1 + \bar{p}q_0, \quad (3.9)$$

where the complementary solution $\bar{p}q_0$ must be derived from the equation (3.8) with no body forces.

4. TRANSFORMATION OF STRESS COMPONENTS

Consider a change to axes On, Os obtained by rotation of Ox, Oy through an angle α (see figure 1), and let \mathbf{n}, \mathbf{s} be unit vectors parallel to the new axes. Then $\mathbf{n} = e^{i\alpha}\mathbf{i}$, and $\mathbf{s} = e^{i\beta}\mathbf{j}$, where $\beta = \frac{1}{2}\pi + \alpha$, and if $\mathbf{R}_n, \mathbf{R}_s, \mathbf{R}_x, \mathbf{R}_y$, are the stress vectors across elements perpendicular to $\mathbf{n}, \mathbf{s}, \mathbf{i}, \mathbf{j}$, respectively, we have, from the equilibrium of the infinitesimal wedges of figures 2 and 3, that

$$\mathbf{R}_n = \cos \alpha \mathbf{R}_x + \sin \alpha \mathbf{R}_y, \quad \mathbf{R}_s = \cos \beta \mathbf{R}_x + \sin \beta \mathbf{R}_y. \quad (4.1)$$

$$\text{But} \quad \mathbf{R}_x = \bar{x}\bar{x}\mathbf{i} + \bar{x}\bar{y}\mathbf{j} = \{\bar{x}\bar{x} + i\bar{x}\bar{y}\}\mathbf{i}, \quad \mathbf{R}_y = \{\bar{x}\bar{y} + i\bar{y}\bar{y}\}\mathbf{j}. \quad (4.2)$$

$$\text{Also} \quad \mathbf{R}_n = \bar{n}\bar{n}\mathbf{n} + \bar{s}\bar{s}\mathbf{s} = \{\bar{n}\bar{n} + i\bar{s}\bar{n}\}\mathbf{n}$$

$$\text{or} \quad \mathbf{R}_n = \{\bar{n}\bar{n} + i\bar{s}\bar{n}\}e^{i\alpha}\mathbf{i}, \quad \mathbf{R}_s = \{\bar{s}\bar{s} + i\bar{s}\bar{n}\}e^{i\alpha}\mathbf{i}. \quad (4.3)$$

From (4.1)–(4.3) we find

$$\begin{aligned} \{\bar{n}\bar{n} + i\bar{s}\bar{n}\}e^{i\alpha} &= \cos \alpha \{\bar{x}\bar{x} + i\bar{x}\bar{y}\} + \sin \alpha \{\bar{x}\bar{y} + i\bar{y}\bar{y}\}, \\ \{\bar{s}\bar{s} + i\bar{s}\bar{n}\}e^{i\alpha} &= -\sin \alpha \{\bar{x}\bar{x} + i\bar{x}\bar{y}\} + \cos \alpha \{\bar{x}\bar{y} + i\bar{y}\bar{y}\}, \end{aligned}$$

and adding and subtracting these equations in turn, and dividing by $e^{i\alpha}$, we have

$$\widehat{nn} + \widehat{ss} = \widehat{xx} + \widehat{yy}, \quad (4.4)$$

$$\widehat{nn} - \widehat{ss} + 2i\widehat{sn} = e^{-2i\alpha}(\widehat{xx} - \widehat{yy} + 2i\widehat{xy}). \quad (4.5)$$

The result (4.4) is well known and widely used; on the other hand, the almost equally simple equation of transformation (4.5) does not seem to be known or at any rate *used* in this form.* Its usefulness may already be realized from equations (2.5) and (3.8). These stress combinations turn out to be all important in a complex variable treatment of two-dimensional elasticity, and their consistent employment will be found to be a large factor in the algebraic simplicity of the method of complex potentials.

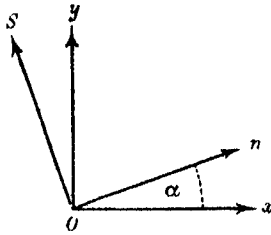


FIGURE 1

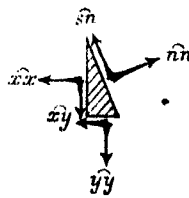


FIGURE 2

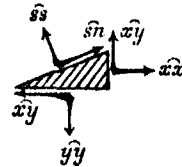


FIGURE 3

We next deal with two elementary states of stress for use later on. Consider the state of stress represented by

$$\widehat{xx} = \widehat{yy} = T, \quad \widehat{xy} = 0. \quad (4.6)$$

A change to parallel axes through any point of the plane does not affect these stresses, and if we choose as origin any point on a curve in the plane whose normal at the point makes an angle α with the x -axis, and then change axes so that On , Os are the normal and tangent respectively, (4.4) and (4.5) at once give

$$\widehat{nn} = \widehat{ss} = T, \quad \widehat{sn} = 0, \quad (4.7)$$

* The following mnemonic symbolism has been brought to the writer's notice by H. G. Hopkins and by Professor L. M. Milne-Thomson.

Let z^2 (i.e. xx), xy , y^2 denote \widehat{xx} , \widehat{xy} , \widehat{yy} respectively, then

$$z\bar{z} = x^2 + y^2 \quad \text{and} \quad z^2 = x^2 - y^2 + 2ixy$$

and so denote $\widehat{xx} + \widehat{yy}$ and $\widehat{xx} - \widehat{yy} + 2i\widehat{xy}$ respectively, and if we change to new axes by turning through an angle α , so that $z' = n + is = ze^{-i\alpha}$, then $z'\bar{z}'$ and z'^2 denote $\widehat{nn} + \widehat{ss}$ and $\widehat{nn} - \widehat{ss} + 2i\widehat{sn}$ respectively. But $z'\bar{z}' = z\bar{z}$ and $z'^2 = z^2 e^{-2i\alpha}$, so that

$$\widehat{nn} + \widehat{ss} = \widehat{xx} + \widehat{yy} \quad \text{and} \quad \widehat{nn} - \widehat{ss} + 2i\widehat{sn} = e^{-2i\alpha}(\widehat{xx} - \widehat{yy} + 2i\widehat{xy}).$$

While not a proof, this is a quick means of recovering (4.4) and (4.5).

and since this holds at all points on any such curve, any curved boundary is only acted upon by an all-round tension $\widehat{nn} = T$.

Again, if the material is in a state of simple tension in a direction making an angle α with the x -axis, so that

$$\widehat{nn} = T, \quad \widehat{ss} = 0, \quad \widehat{sn} = 0, \quad (4.8)$$

equations (4.4) and (4.5) at once give the results

$$\widehat{xx} + \widehat{yy} = T, \quad \widehat{xx} - \widehat{yy} + 2i\widehat{xy} = Te^{2i\alpha}, \quad (4.9)$$

which we shall find useful in this combined form later on. Separation gives

$$\widehat{xx} = T \cos^2 \alpha, \quad \widehat{yy} = T \sin^2 \alpha, \quad \widehat{xy} = T \sin \alpha \cos \alpha, \quad (4.10)$$

and analyses the simple tension in a direction α into simple tensions $T \cos^2 \alpha$, $T \sin^2 \alpha$ parallel to the axes, together with a shear $T \sin \alpha \cos \alpha$.

5. CURVILINEAR CO-ORDINATES

We take curvilinear co-ordinates ξ, η defined by

$$z = f(\zeta), \quad z = x + iy, \quad \zeta = \xi + i\eta, \quad (5.1)$$

so that*

$$dz = f'(\zeta) d\zeta = J e^{i\alpha} d\zeta, \quad (5.2)$$

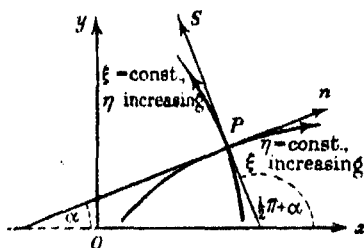


FIGURE 4

where J and α are real, so that

$$J^2 = \frac{d\zeta}{dz} \frac{d\bar{\zeta}}{d\bar{z}}, \quad e^{-2i\alpha} = \frac{d\zeta}{dz} \frac{d\bar{z}}{d\bar{\zeta}}. \quad (5.3)$$

Consider increments dz_1, dz_2 in z along the curves $\eta = \text{const.}$, $\xi = \text{const.}$ at P (see figure 4). Then

$$dz_1 = J e^{i\alpha} d\xi, \quad dz_2 = J e^{i\alpha} i d\eta = J e^{i(\frac{1}{2}\pi + \alpha)} d\eta,$$

and since increments of position vector r are given by $dr = dz$ and $\arg dz_2 = \arg dz_1 + \frac{1}{2}\pi$, the curves are clearly orthogonal. Hence we may

change to axes at P where the directions ξ increasing, η increasing respectively can be taken as the axes On , Os of the last section. Replacing n by ξ , s by η , and using (5.3), equations (4.4) and (4.5) become

$$\xi\xi + \eta\eta = \widehat{x}\widehat{x} + \widehat{y}\widehat{y}, \quad (5.4)$$

$$\xi\xi - \eta\eta + 2i\xi\eta = \frac{d\xi}{dz} \frac{d\bar{z}}{d\bar{\xi}} (\widehat{x}\widehat{x} - \widehat{y}\widehat{y} + 2i\widehat{x}\widehat{y}). \quad (5.5)$$

These, used with (5.1), are the simple formulae for the transformation of stress in two-dimensional orthogonal curvilinear co-ordinates.

6. THE COMPLEX POTENTIAL FUNCTIONS

Now it is clear that the complex body-stress equation (3.8), in the two cases of plane strain and generalized plane stress, is satisfied by the introduction of a new function $F(z, \bar{z})$ of the co-ordinates z, \bar{z} in terms of which

$$\widehat{x}\widehat{x} + \widehat{y}\widehat{y} - 2\rho U = \frac{\partial F(z, \bar{z})}{\partial z}, \quad (6.1)$$

$$\widehat{x}\widehat{x} - \widehat{y}\widehat{y} + 2i\widehat{x}\widehat{y} = -\frac{\partial F(z, \bar{z})}{\partial \bar{z}}. \quad (6.2)$$

From (2.5) and (6.2) we find

$$4\mu \frac{\partial D}{\partial \bar{z}} = -\frac{\partial F(z, \bar{z})}{\partial \bar{z}},$$

whence

$$4\mu D = f(z) - F(z, \bar{z}), \quad (6.3)$$

where $f(z)$ is a function of a *single* complex variable z . Also from (1.5) and (2.5) we have

$$4(\lambda + \mu) \frac{\partial D}{\partial z} = \widehat{x}\widehat{x} + \widehat{y}\widehat{y} + 4i(\lambda + \mu)\varpi. \quad (6.4)$$

Hence from (6.1), (6.3) and (6.4), writing

$$f(z) = 2(1 - \eta)\Omega(z), \quad (6.5)$$

where, in plane strain η is Poisson's ratio defined by

$$\eta = \lambda/2(\lambda + \mu), \quad (6.6)$$

and in generalized plane stress must be replaced by the modified Poisson's ratio σ obtained by replacing λ by λ' of (1.6) and related to Poisson's ratio η by the equation

$$(1 - \sigma)(1 + \eta) = 1, \quad (6.7)$$

we find

$$\widehat{x}\widehat{x} + \widehat{y}\widehat{y} - (\rho U - 2\mu i\varpi)/(1 - \eta) = \Omega'(z), \quad (6.8)$$

with the usual notation for the differentiation of a function of a single variable with respect to the argument. There will be no occasion to confuse the elastic constant η with the curvilinear co-ordinate η .

Now, from (6.1), (6.5), (6.6) and (6.8), using (3.2), we have, for plane strain,

$$2 \frac{\partial F(z, \bar{z})}{\partial z} = \Omega'(z) + \bar{\Omega}'(\bar{z}) - 2\rho \left(\frac{1-2\eta}{1-\eta} \right) \frac{\partial W(z, \bar{z})}{\partial z},$$

where $\bar{\Omega}(\)$ is derived from $\Omega(\)$ by changing i to $-i$ wherever it occurs in the definition of $\Omega(\)$. On integration with respect to z we may write

$$2F(z, \bar{z}) = \Omega(z) + z\bar{\Omega}'(\bar{z}) - \gamma\rho W(z, \bar{z}) + \bar{\omega}'(\bar{z}), \quad (6.9)$$

where $\omega(z)$ is another function of the single complex variable, and γ is given by

$$\gamma = 2(1-2\eta)/(1-\eta) \quad \text{or} \quad \gamma = 2(1-\eta), \quad (6.10)$$

according as we are dealing with plane strain or generalized plane stress.

Hence (6.3), (6.5) and (6.9) give the solution for the displacements in terms of the two complex potential functions $\Omega(z)$, $\omega(z)$ as

$$8\mu(u + iv) = \kappa\Omega(z) - z\bar{\Omega}'(\bar{z}) - \bar{\omega}'(\bar{z}) + \gamma\rho W(z, \bar{z}), \quad (6.11)$$

where κ is an elastic constant, in terms of which γ is always given by

$$\gamma = 4(\kappa - 1)/(\kappa + 1), \quad (6.12)$$

$$\text{and we have } (1 + \kappa)/(1 - \eta) = 4 \quad \text{or} \quad (1 + \kappa)(1 + \eta) = 4, \quad (6.13)$$

according as we are dealing with plane strain or generalized plane stress.

From (6.1), (6.2), (6.9) and (6.12), the stresses are given in terms of the complex potentials and the body-force function $W(z, \bar{z})$ by the equations

$$\widehat{x}\widehat{x} + \widehat{y}\widehat{y} = \frac{1}{2} \left\{ \Omega'(z) + \bar{\Omega}'(\bar{z}) + \frac{8\rho}{\kappa + 1} \frac{\partial W(z, \bar{z})}{\partial z} \right\}, \quad (6.14)$$

$$\widehat{x}\widehat{x} - \widehat{y}\widehat{y} + 2i\widehat{x}\widehat{y} = -\frac{1}{2} \left\{ z\bar{\Omega}''(\bar{z}) + \bar{\omega}''(\bar{z}) - \gamma\rho \frac{\partial W(z, \bar{z})}{\partial \bar{z}} \right\}. \quad (6.15)$$

We can readily find the stresses and displacements in terms of curvilinear co-ordinates. For if u_ξ , u_η are the displacements at P (figure 4) resolved along the directions ξ , η increasing respectively, we have

$$(u + iv)\mathbf{i} = \mathbf{D} = (u_\xi + iu_\eta)\mathbf{n} = (u_\xi + iu_\eta)e^{i\alpha}\mathbf{i},$$

so that, from (5.2) and (6.11) we have

$$8\mu(u_\xi + iu_\eta) = \frac{1}{J} \frac{d\bar{z}}{d\xi} \{ \kappa\Omega(z) - z\bar{\Omega}'(\bar{z}) - \bar{\omega}'(\bar{z}) + \gamma\rho W(z, \bar{z}) \}, \quad (6.16)$$

whilst (5.4), (5.5), (6.14) and (6.15) give

$$\xi\xi + \bar{\eta}\bar{\eta} = \frac{1}{2} \left\{ \Omega'(z) + \bar{\Omega}'(\bar{z}) + \frac{8\rho}{\kappa+1} \frac{\partial W(z, \bar{z})}{\partial z} \right\}, \quad (6.17)$$

$$\xi\xi - \bar{\eta}\bar{\eta} + 2i\xi\bar{\eta} = -\frac{1}{2} \frac{d\xi}{dz} \frac{d\bar{z}}{d\bar{\xi}} \left\{ z\bar{\Omega}''(\bar{z}) + \bar{\omega}''(\bar{z}) - \gamma\rho \frac{\partial W(z, \bar{z})}{\partial \bar{z}} \right\}, \quad (6.18)$$

which latter can be written

$$\xi\xi - \bar{\eta}\bar{\eta} + 2i\xi\bar{\eta} = -\frac{1}{2} \frac{d\xi}{dz} \frac{d}{d\bar{\xi}} \{ z\bar{\Omega}'(\bar{z}) + \bar{\omega}'(\bar{z}) - \gamma\rho W(z, \bar{z}) \}. \quad (6.19)$$

In these formulae we suppose z and \bar{z} to be expressed in terms of $\xi, \bar{\xi}$ by means of (5.1).

7. RESTRICTIONS ON THE COMPLEX POTENTIALS

There will be certain restrictions on the nature of the functions $\Omega(z)$, $\omega(z)$. Clearly the stresses must be single valued; on the other hand certain many-valued displacements are capable of physical interpretations as *dislocations*, which can be realized in a two-dimensional region of multiple connectivity by drawing a curve which reduces the connectivity when the material is cut along this curve, displacing the material so that the curve is given a small rigid body displacement, the material between the two positions of the curve being cut away and the edges then recemented (e.g. see Coker and Filon 1931, p. 503). If we start from a point $P(z, \bar{z})$ in the elastic material and return always to P via a path in the material, the change in a function ϕ must have the same value for all circuits reducible to one another and must vanish if ϕ is single-valued. Following Filon, we term this change the cyclic function of ϕ for the circuit typical of a given particular connectivity and write it $Cy\phi$. It is readily shown (e.g. see Coker and Filon 1931, p. 513) that Cy and $\partial/\partial z$, $\partial/\partial \bar{z}$ are commutative symbols of operation.

When there are no body forces the conditions that the stresses shall be single-valued are, from (6.14) and (6.15),

$$Cy \{ \Omega'(z) + \bar{\Omega}'(\bar{z}) \} = 0, \quad (7.1)$$

and

$$Cy \{ z\bar{\Omega}''(\bar{z}) + \bar{\omega}''(\bar{z}) \} = 0,$$

which gives

$$Cy \bar{\omega}''(\bar{z}) = 0, \quad Cy \bar{\Omega}''(\bar{z}) = 0 \quad (7.2)$$

separately.

Further, since the stresses must be single-valued, it is clear that if $Cy(u+iv) = U+iV$, then U, V are displacements due to a system of zero

stresses, so that in fact they form a two-dimensional rigid body displacement, i.e. we may write

$$Cy(u + iv) = \alpha + i\beta z, \quad (7.3)$$

where α is a small complex constant representing a small displacement of translation, and β is a small real constant, the corresponding terms representing a small rigid body displacement of rotation of angle β . Hence from (6.11)

$$Cy\{\kappa\Omega(z) - z\bar{\Omega}'(\bar{z}) - \bar{\omega}'(\bar{z})\} = 8\mu(\alpha + i\beta z), \quad (7.4)$$

and differentiation with respect to z and \bar{z} respectively gives

$$Cy\{\kappa\Omega'(z) - \bar{\Omega}'(\bar{z})\} = 8\mu i\beta \quad (7.5)$$

and equations (7.2) once more. From (7.1) and (7.5), we find

$$Cy\Omega'(z) = 8\mu i\beta/(\kappa + 1), \quad (7.6)$$

which is clearly consistent with (7.1) and (7.2). So far, therefore, the restrictions upon the complex potentials $\Omega(z)$, $\omega(z)$ are given by (7.2) and (7.6).

When the body forces are of the type given by (3.4), it is clear, since $\partial W/\partial z$ and $\partial W/\partial \bar{z}$ are single-valued, that the restrictions upon $\Omega(z)$ and $\omega(z)$ are, so far, the same as in the case of no body forces. To understand fully the nature of $Cy\Omega(z)$ and $Cy\omega(z)$, we must next consider the stress resultants round a circuit.

8. THE STRESS RESULTANTS ROUND A CIRCUIT

For an element ds of a circuit C we have

$$dz\mathbf{i} = d\mathbf{s} = ds e^{i(\pi+\alpha)}\mathbf{i},$$

$$\text{whence} \quad \frac{dz}{ds} = ie^{i\alpha}, \quad \frac{dz}{ds} \frac{d\bar{z}}{ds} = 1. \quad (8.1)$$

The stress vector \mathbf{R}_n across the element is

$$\mathbf{R}_n = \widehat{x}\mathbf{n}\mathbf{i} + \widehat{y}\mathbf{n}\mathbf{j} = \{\widehat{x}\mathbf{n} + i\widehat{y}\mathbf{n}\}\mathbf{i}; \quad (8.2)$$

$$\text{hence from (4.3)} \quad \widehat{x}\mathbf{n} + i\widehat{y}\mathbf{n} = \{\widehat{n}\mathbf{n} + i\widehat{s}\mathbf{n}\}e^{i\alpha}. \quad (8.3)$$

Now the force and couple resultants $\mathbf{R} = \{X + iY\}\mathbf{i}$, $\mathbf{M} = M\mathbf{k}$ at the origin $z = 0$ for the stress system exerted upon the material within the closed boundary C are

$$\mathbf{R} = \int_C \mathbf{R}_n ds, \quad \mathbf{M} = \int_C [\mathbf{r}, \mathbf{R}_n] ds,$$

where $r = zi$. Hence, making use of (8.1) and (8.3), we have

$$X + iY = \int_C (\bar{x}\bar{n} + i\bar{y}\bar{n}) ds = -i \int_C (\bar{n}\bar{n} + i\bar{s}\bar{n}) \frac{dz}{ds} ds, \quad (8.4)$$

$$M = \int_C (x\bar{y}\bar{n} - y\bar{x}\bar{n}) ds = \text{real part of } -i \int_C \bar{z}(\bar{x}\bar{n} + i\bar{y}\bar{n}) ds,$$

or we may write $M + iM^* = -i \int_C \bar{z}(\bar{x}\bar{n} + i\bar{y}\bar{n}) \frac{dz}{ds} ds. \quad (8.5)$

Now, from (4.4) and (4.5)

$$4(\bar{n}\bar{n} + i\bar{s}\bar{n}) = 2(\bar{x}\bar{x} + \bar{y}\bar{y}) + 2(\bar{x}\bar{x} - \bar{y}\bar{y} + 2i\bar{x}\bar{y})e^{-2i\alpha},$$

and so, in terms of the complex potentials from (6.14) and (6.15), using (8.1)

$$4(\bar{n}\bar{n} + i\bar{s}\bar{n}) = \Omega'(z) + \bar{\Omega}'(\bar{z}) + \left(\frac{dz}{ds}\right)^2 \left\{ z\bar{\Omega}''(\bar{z}) + \bar{\omega}''(z) - \gamma\rho \frac{\partial W}{\partial \bar{z}} \right\} + \frac{8\rho}{\kappa+1} \frac{\partial W}{\partial z}.$$

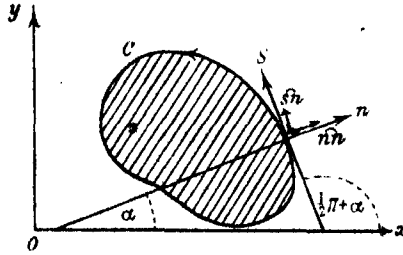


FIGURE 5

Hence from (8.4), again using (8.1), also (6.12),

$$4i(X + iY) = \int_C \left\{ \Omega'(z) + \bar{\Omega}'(\bar{z}) \right\} \frac{dz}{ds} ds + \int_C \left\{ \frac{8\rho}{\kappa+1} \frac{dW(z, \bar{z})}{ds} + \frac{d\bar{\omega}'(z)}{ds} + z \frac{d\bar{\Omega}'(\bar{z})}{ds} \right\} ds - 4\rho \int \frac{\partial W(z, \bar{z})}{\partial \bar{z}} \frac{d\bar{z}}{ds} ds,$$

whence

$$X + iY = -\frac{1}{4}i \int_C \frac{d}{ds} \left\{ \Omega(z) + z\bar{\Omega}'(\bar{z}) + \bar{\omega}'(z) + \frac{8\rho}{\kappa+1} W(z, \bar{z}) \right\} ds + i\rho \int_C \frac{\partial W(z, \bar{z})}{\partial \bar{z}} \frac{d\bar{z}}{ds} ds,$$

or with the body forces of type given by (3.4), we have

$$X + iY = -\frac{1}{4}iCy \left\{ \Omega(z) + z\bar{\Omega}'(\bar{z}) + \bar{\omega}'(z) \right\} + i\rho \int_C (\bar{c}z + \frac{1}{2}dz^2) d\bar{z}.$$

Now, converting line integrals round the bounding curve C to integrals over the area S enclosed by C , we find

$$\int_C z d\bar{z} = -2iS, \quad \int_C z^2 d\bar{z} = -4i \int_C z dS = -4iz_0 S, \quad (8.6)$$

where the point $z = z_0$ is the centroid of the area S . Hence, finally,

$$X + iY = -\frac{1}{4}Cy \{ \Omega(z) + z\bar{\Omega}'(\bar{z}) + \bar{\omega}'(\bar{z}) \} + 2\rho S(\bar{c} + dz_0). \quad (8.7)$$

In similar fashion (8.5) becomes

$$\begin{aligned} 4(M + iM^*) &= - \int_C \left\{ \bar{z} \frac{d\Omega(z)}{ds} + \bar{z} \frac{dz}{ds} \bar{\Omega}'(\bar{z}) + \bar{z} \frac{d\bar{\omega}'(\bar{z})}{ds} + z\bar{z} \frac{d\bar{\Omega}'(\bar{z})}{ds} + \frac{8\rho}{\kappa+1} \bar{z} \frac{dW(z, \bar{z})}{ds} \right\} ds \\ &\quad + 4\rho \int_C \frac{\partial W(z, \bar{z})}{\partial \bar{z}} \frac{d\bar{z}}{ds} ds \\ &= - \int_C \frac{d}{ds} \left\{ z\bar{z} \bar{\Omega}'(\bar{z}) + z\bar{\omega}'(\bar{z}) - \bar{\omega}(\bar{z}) + \frac{8\rho}{\kappa+1} \bar{z} W(z, \bar{z}) \right\} ds + \frac{8\rho}{\kappa+1} \int_C W(z, \bar{z}) \frac{d\bar{z}}{ds} ds \\ &\quad + 4\rho \int_C \bar{z} \frac{d\bar{z}}{ds} \frac{\partial W(z, \bar{z})}{\partial \bar{z}} ds + \int_C \left\{ z \frac{d\bar{\Omega}'(\bar{z})}{ds} - \bar{z} \frac{d\Omega(z)}{ds} \right\} ds. \end{aligned}$$

The last integral can be dropped without affecting M , since the integrand is a pure imaginary. Using body forces given by (3.4), we have

$$\begin{aligned} M + iM^* &= -\frac{1}{4}Cy \{ z\bar{z} \bar{\Omega}'(\bar{z}) + z\bar{\omega}'(\bar{z}) - \bar{\omega}(\bar{z}) \} + \frac{4\rho c}{\kappa+1} \int_C z^2 d\bar{z} \\ &\quad + 2\rho \left(\frac{\kappa+3}{\kappa+1} \right) \left\{ 2\bar{c} \int_C z\bar{z} d\bar{z} + d \int_C z^2 \bar{z} d\bar{z} \right\}. \end{aligned}$$

Again converting line integrals to surface integrals we find

$$\int_C z\bar{z} d\bar{z} = -2i\bar{z}_0 S, \quad \int_C z^2 \bar{z} d\bar{z} = -4i \int_S z\bar{z} dS = -4iI, \quad (8.8)$$

where I is the second moment of the area S about the axis through $z = 0$ perpendicular to the area S . But d is wholly real in the cases considered and from (7.6) $Cy \{ z\bar{z} \bar{\Omega}'(\bar{z}) \}$ is a pure imaginary, so that again modifying the meaning of M^* without affecting M , we may write

$$M + iM^* = \frac{1}{4}Cy \{ \bar{\omega}(z) - \bar{z}\bar{\omega}'(\bar{z}) \} - \frac{8\rho i S}{\kappa+1} \{ 2cz_0 + (\kappa+3)\bar{c}\bar{z}_0 \},$$

or equally well, changing the sign of i ,

$$M = \text{real part of } \frac{1}{4}Cy \{ \omega(z) - z\omega'(z) \} + \frac{8\rho i S}{\kappa+1} \{ 2\bar{c}\bar{z}_0 + (\kappa+3)cz_0 \}. \quad (8.9)$$

The results (7.2), (7.4), (8.7), (8.9) concerning the cyclic functions of the two complex potentials are immediately useful in particular problems as will be seen in the later sections.

9. EFFECT ON THE COMPLEX POTENTIALS OF CHANGE OF ORIGIN

Since the stresses cannot change with a change of origin, it follows from (6.14) and (6.15) in the case of no body forces that $\Omega'(z)$ and $\bar{z}\Omega''(z) + \omega''(z)$ must be independent of the choice of origin. Hence, if $\Omega_C(z_1)$, $\omega_C(z_1)$ are the complex potentials appropriate to the origin at C ($z = c$, $z_1 = 0$), then

$$\Omega'(z) = \Omega'_C(z_1),$$

$$\text{and} \quad \bar{z}\Omega''(z) + \omega''(z) = (\bar{z} - \bar{c})\Omega''_C(z_1) + \omega''_C(z_1),$$

$$\text{or} \quad \omega''_C(z_1) = \omega''(z) + \bar{c}\Omega''(z).$$

Hence we may write

$$\Omega_C(z_1) = \Omega(z), \quad \omega_C(z_1) = \omega(z) + \bar{c}\Omega(z), \quad (9.1)$$

and these will be seen later to prove helpful in combining elementary solutions referred to different origins.

The displacements must also be independent of the choice of origin, and the relations (9.1) are readily seen to satisfy (6.11) in this respect.

10. RELATION TO OTHER METHODS

When there are no body forces let us write

$$\chi + i\chi^* = \frac{1}{2}\{z\Omega(z) + \omega(z)\}, \quad (10.1)$$

$$\text{whence} \quad 4\frac{\partial^2\chi}{\partial z^2} = \frac{1}{2}\{z\bar{\Omega}''(\bar{z}) + \bar{\omega}''(\bar{z})\}, \quad 4\frac{\partial^2\chi}{\partial z\partial\bar{z}} = \frac{1}{2}\{\Omega'(z) + \bar{\Omega}'(\bar{z})\}.$$

Hence, using (2.4), (6.14) and (6.15) become

$$\widehat{xx} - \widehat{yy} + 2i\widehat{xy} = -4\frac{\partial^2\chi}{\partial\bar{z}^2} = -\left\{\frac{\partial^2\chi}{\partial x^2} - \frac{\partial^2\chi}{\partial y^2} + 2i\frac{\partial^2\chi}{\partial x\partial y}\right\},$$

$$\widehat{xx} + \widehat{yy} = 4\frac{\partial^2\chi}{\partial z\partial\bar{z}} = \frac{\partial^2\chi}{\partial x^2} + \frac{\partial^2\chi}{\partial y^2},$$

and separation gives

$$\widehat{xx} = \frac{\partial^2\chi}{\partial y^2}, \quad \widehat{yy} = \frac{\partial^2\chi}{\partial x^2}, \quad \widehat{xy} = -\frac{\partial^2\chi}{\partial x\partial y}, \quad (10.2)$$

so that, when there is no body force, we have identified Airy's stress function χ with the real part of $\frac{1}{2}\{\bar{z}\Omega(z) + \omega(z)\}$.

$$\text{If we write} \quad \Omega(z) = \Phi + i\Psi, \quad \omega(z) = \phi + i\psi, \quad (10.3)$$

we find, using (2.3) and the Cauchy-Riemann relations, that

$$\begin{aligned} \Omega'(z) &= \frac{\partial\Phi}{\partial x} - i\frac{\partial\Phi}{\partial y} = \frac{\partial\Psi}{\partial y} + i\frac{\partial\Psi}{\partial x}, \\ \bar{z}\Omega'(z) &= x\frac{\partial\Phi}{\partial x} - y\frac{\partial\Phi}{\partial y} + i\left(x\frac{\partial\Psi}{\partial x} - y\frac{\partial\Psi}{\partial y}\right), \\ \omega'(z) &= \frac{\partial\phi}{\partial x} - i\frac{\partial\phi}{\partial y}. \end{aligned}$$

Hence (6.11) gives, when there are no body forces,

$$8\mu u = \kappa\Phi - \left(x\frac{\partial\Phi}{\partial x} - y\frac{\partial\Phi}{\partial y}\right) - \frac{\partial\phi}{\partial x}, \quad (10.4)$$

$$8\mu v = \kappa\Psi + \left(x\frac{\partial\Psi}{\partial x} - y\frac{\partial\Psi}{\partial y}\right) - \frac{\partial\phi}{\partial y}. \quad (10.5)$$

These show that Φ , Ψ , ϕ are merely simple multiples of plane harmonic functions used by Bricas (1937, p. 75), and that in terms of these functions Airy's stress function is given from (10.1) as

$$\chi = \frac{1}{2}(x\Phi + y\Psi + \phi). \quad (10.6)$$

$$\text{Again, if we write} \quad \Theta = \frac{\partial\Phi}{\partial x} = \frac{\partial\Psi}{\partial y} = \frac{1}{2}\{\Omega'(z) + \bar{\Omega}'(\bar{z})\}, \quad (10.7)$$

$$\begin{aligned} \text{then} \quad z\bar{\Omega}'(\bar{z}) &= 2z\frac{\partial\Theta}{\partial\bar{z}} = (x+iy)\left(\frac{\partial\Theta}{\partial x} + i\frac{\partial\Theta}{\partial y}\right) \\ &= x\frac{\partial\Theta}{\partial x} - y\frac{\partial\Theta}{\partial y} + i\left(y\frac{\partial\Theta}{\partial x} + x\frac{\partial\Theta}{\partial y}\right). \end{aligned}$$

$$\begin{aligned} \text{Also} \quad \bar{\omega}''(\bar{z}) &= \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\left(\frac{\partial\phi}{\partial x} + i\frac{\partial\phi}{\partial y}\right) \\ &= \frac{1}{2}\left\{\frac{\partial^2\phi}{\partial x^2} - \frac{\partial^2\phi}{\partial y^2} + 2i\frac{\partial^2\phi}{\partial x\partial y}\right\} \\ &= \frac{\partial^2\phi}{\partial x^2} + i\frac{\partial^2\phi}{\partial x\partial y} = -\frac{\partial^2\phi}{\partial y^2} + i\frac{\partial^2\phi}{\partial x\partial y}. \end{aligned}$$

Hence, in the case of no body forces, (6.14) and (6.15) can be written

$$\widehat{xx} + \widehat{yy} = \Theta,$$

$$\widehat{xx} - \widehat{yy} + 2i\widehat{xy} = -\frac{1}{2} \left\{ x \frac{\partial \Theta}{\partial x} - y \frac{\partial \Theta}{\partial y} + \frac{\partial^2 \phi}{\partial x^2} + i \left(y \frac{\partial \Theta}{\partial x} + x \frac{\partial \Theta}{\partial y} + \frac{\partial^2 \phi}{\partial x \partial y} \right) \right\},$$

$$\text{whence} \quad \widehat{xx} = \frac{1}{2} \left\{ 2\Theta - \left(x \frac{\partial \Theta}{\partial x} - y \frac{\partial \Theta}{\partial y} \right) + \frac{\partial^2 \phi}{\partial y^2} \right\}, \quad (10.8)$$

$$\widehat{yy} = \frac{1}{4} \left\{ 2\Theta + \left(x \frac{\partial \Theta}{\partial x} - y \frac{\partial \Theta}{\partial y} \right) + \frac{\partial^2 \phi}{\partial x^2} \right\}, \quad (10.9)$$

$$\widehat{xy} = -\frac{1}{4} \left\{ x \frac{\partial \Theta}{\partial y} + y \frac{\partial \Theta}{\partial x} + \frac{\partial^2 \phi}{\partial x \partial y} \right\}, \quad (10.10)$$

giving the stresses in terms of two plane harmonic functions Θ , ϕ , which are simple multiples of plane harmonic functions used by Bricas (1932, p. 11; 1937, p. 76).

In curvilinear co-ordinates, still taking the case of no body force, we have

$$\widehat{\xi\xi} + \widehat{\eta\eta} = \widehat{xx} + \widehat{yy} = \Theta, \quad (10.11)$$

and, writing $z\bar{z} = r^2$, $J^{-1} = h$, (6.19) becomes, changing the sign of i through-out,

$$\widehat{\xi\xi} - \widehat{\eta\eta} - 2i\widehat{\xi\eta} = -\frac{1}{2} h^2 \frac{\partial r^2}{\partial \xi} \frac{d\Omega'(z)}{d\xi} - \frac{1}{2} \frac{\partial}{\partial \xi} \left\{ h^2 \frac{d\omega(z)}{d\xi} \right\}, \quad (10.12)$$

$$\begin{aligned} \text{Now} \quad 2 \frac{\partial r^2}{\partial \xi} \frac{d\Omega'(z)}{d\xi} &= \left(\frac{\partial r^2}{\partial \xi} - i \frac{\partial r^2}{\partial \eta} \right) \left(\frac{\partial \Theta}{\partial \xi} - i \frac{\partial \Theta}{\partial \eta} \right) \\ &= [r^2, \Theta] - i(r^2, \Theta), \end{aligned}$$

where $[U, V]$, (U, V) are abbreviations given by

$$[U, V] = \frac{\partial U}{\partial \xi} \frac{\partial V}{\partial \xi} - \frac{\partial U}{\partial \eta} \frac{\partial V}{\partial \eta}, \quad (U, V) = \frac{\partial U}{\partial \xi} \frac{\partial V}{\partial \eta} + \frac{\partial U}{\partial \eta} \frac{\partial V}{\partial \xi}. \quad (10.13)$$

$$\begin{aligned} \text{Again} \quad 2 \frac{\partial}{\partial \xi} \left\{ h^2 \frac{d\omega(z)}{d\xi} \right\} &= \left(\frac{\partial}{\partial \xi} - i \frac{\partial}{\partial \eta} \right) \left\{ h^2 \left(\frac{\partial \phi}{\partial \xi} - i \frac{\partial \phi}{\partial \eta} \right) \right\} \\ &= [h^2, \phi] - i(h^2, \phi) + h^2 \left\{ \frac{\partial^2 \phi}{\partial \xi^2} - \frac{\partial^2 \phi}{\partial \eta^2} - 2i \frac{\partial^2 \phi}{\partial \xi \partial \eta} \right\}. \end{aligned}$$

Hence from (10.11) and (10.12) separation of the stresses gives

$$\xi\xi = \frac{1}{2}\Theta - \frac{1}{8}\left\{h^2[r^2, \Theta] + [h^2, \phi] + 2h^2\frac{\partial^2\phi}{\partial\xi^2}\right\}, \quad (10.14)$$

$$\eta\eta = \frac{1}{2}\Theta + \frac{1}{8}\left\{h^2[r^2, \Theta] + [h^2, \phi] - 2h^2\frac{\partial^2\phi}{\partial\eta^2}\right\}, \quad (10.15)$$

$$\xi\eta = -\frac{1}{8}\left\{h^2(r^2, \Theta) + (h^2, \phi) + 2h^2\frac{\partial^2\phi}{\partial\xi\partial\eta}\right\}, \quad (10.16)$$

which are equivalent to the results obtained by Bricas (1937, p. 98).

It cannot be too strongly emphasized here, however, that the resolution of the complex potentials into such plane harmonic functions as Φ , Ψ , ϕ , ψ and the use of Θ is entirely foreign to the spirit of the method proposed in this paper. Such resolution is entirely unnecessary, and leads to almost as much labour in fitting solutions as does the Airy stress function method, although, in side-tracking the biharmonic equation, Bricas's method has some advantages over the Airy stress function method. In the method of this paper, a single complex boundary condition replaces the double condition of the usual methods, and this is indeed a guide to the form of the solutions required, whilst separation of reals and imaginaries is deferred until the final stages of a problem with a very marked economy of effort. It should perhaps be mentioned here that when the present writer first developed these methods, he was unaware of the interesting plane harmonic methods of M. Bricas.

11. SOLUTIONS IN CARTESIAN AND POLAR CO-ORDINATES

Consider the transformation of co-ordinates given by

$$\zeta = \log z = \log r + i\theta, \quad (11.1)$$

which makes
$$\frac{d\zeta}{dz} = \frac{1}{z}, \quad J = r. \quad (11.2)$$

Equations (6.16)–(6.18) become, with the body forces of (3.4),

$$8\mu(u_r + iu_\theta) = e^{-i\theta}\{\kappa\Omega(z) - z\bar{\Omega}'(\bar{z}) - \bar{\omega}'(\bar{z}) + \frac{1}{2}\gamma\rho(cz^2 + 2\bar{c}z\bar{z} + dz^2\bar{z})\}, \quad (11.3)$$

$$\bar{r}\bar{r} + \bar{\theta}\bar{\theta} = \frac{1}{2}\left\{\Omega'(z) + \bar{\Omega}'(\bar{z}) + \frac{8\rho}{\kappa+1}(cz + \bar{c}\bar{z} + dz\bar{z})\right\}, \quad (11.4)$$

$$\bar{r}\bar{r} - \bar{\theta}\bar{\theta} + 2i\bar{r}\bar{\theta} = -\frac{1}{2}\left\{z\bar{\Omega}''(\bar{z}) + \frac{\bar{z}}{z}\bar{\omega}''(\bar{z}) - \gamma\rho(\bar{c}\bar{z} + \frac{1}{2}dz\bar{z})\right\}. \quad (11.5)$$

Solutions for a concentric ring space

Here we obtain the complex potentials for the following problems relative to the ring space $b \leq r \leq a$:

- (i) equilibrium under uniform shears applied to the boundaries,
- (ii) equilibrium under constant internal and external pressures,
- (iii) equilibrium of a dislocated annulus, under no tractions round the boundaries, so that the strain is due entirely to the dislocation; we shall refer to this as the *pure dislocational* problem of the annulus,
- (iv) steady motion of the annulus with angular velocity n about the centre; also steady rotation of the complete disk bounded by $r = a$.

Of these, the first three are supposed to be under zero body force, the fourth is regarded as a statical problem with the reversed mass accelerations as body forces.

Assuming no body force, the boundary stresses over the circle $|z| = r$ are given from (11.4) and (11.5), following a change of sign of i , as

$$4(\bar{r}\bar{r} - i\bar{r}\bar{\theta}) = \Omega'(z) + \bar{\Omega}'(z) - z\Omega''(z) - \frac{z}{2}\omega''(z). \quad (11.6)$$

Consider first the solution given by the complex potentials

$$\Omega(z) = 0, \quad \omega(z) = C \log z, \quad C = C_1 + iC_2, \quad (11.7)$$

where C_1, C_2 are real. For circuits in the elastic material and embracing the origin

$$Cy \omega(z) = 2\pi i C,$$

so that from (7.4) we see that the solution is non-dislocational in character, whilst from (8.7) and (8.9) it is clear that the stress resultants for the boundary are $X = Y = 0$, $M = -\frac{1}{2}\pi C_2$. Also from (11.6)

$$4(\bar{r}\bar{r} - i\bar{r}\bar{\theta}) = -C/z\bar{z} = -(C_1 + iC_2)/r^2.$$

Hence the complex potentials for the annulus under a uniform positive (anti-clockwise) couple M applied uniformly to the outer boundary, balanced by an equal, opposite and uniformly applied couple M at the inner boundary, are

$$\Omega(z) = 0, \quad \omega(z) = -2i \frac{M}{\pi} \log z. \quad (11.8)$$

The solution given by C_1 in (11.7) can be combined with the elementary solution $\Omega(z) = Az$, $\omega(z) = 0$, where A is real, which gives from (11.6)

$$4(\bar{r}\bar{r} - i\bar{r}\bar{\theta}) = 2A - C_1/r^2.$$

Note that an imaginary constant A merely gives rise to a rigid body displacement of rotation. If now we put $\hat{r} = -p_o$ for $r = a$, and $\hat{r} = -p_i$ for $r = b$, we find on solving for A and C_1 that the solution for the equilibrium of the annulus under internal pressure p_i and external pressure p_o is given by the complex potentials

$$\Omega(z) = -2 \left(\frac{p_o a^2 - p_i b^2}{a^2 - b^2} \right) z, \quad \omega(z) = 4a^2 b^2 \left(\frac{p_i - p_o}{a^2 - b^2} \right) \log z. \quad (11.9)$$

Next consider the complex potentials

$$\Omega(z) = Bz \log z, \quad \omega(z) = 0, \quad (11.10)$$

where B is real. From (8.7) and (8.9) the stress resultants vanish for a circuit embracing the origin, but from (7.4), since $Cy \Omega(z) = 2\pi i Bz$, the solution corresponds to a rotational dislocation of (small) angle β given by

$$B = 4\mu\beta/\pi(\kappa + 1). \quad (11.11)$$

From (11.6) and (11.10) we readily find

$$4(\hat{r} - ir\hat{\theta}) = B(1 + 2 \log r).$$

To obtain the complex potentials for the pure rotational dislocation of the annulus, we need only combine this solution with the previous, taking

$$p_o = \frac{1}{2} B(1 + 2 \log a), \quad p_i = \frac{1}{2} B(1 + 2 \log b),$$

so that, with the value of B given by (11.11), the appropriate complex potentials are

$$\left. \begin{aligned} \Omega(z) &= \frac{4\mu\beta}{\pi(\kappa + 1)} \left\{ z \log z - \left(\frac{1}{2} + \frac{a^2 \log a - b^2 \log b}{a^2 - b^2} \right) z \right\}, \\ \omega(z) &= \frac{8\mu\beta}{\pi(\kappa + 1)} \frac{a^2 b^2 \log(b/a)}{(a^2 - b^2)} \log z. \end{aligned} \right\} \quad (11.12)$$

This well-known problem, first solved by Michell, is often referred to in a less descriptive manner as the pure dislocation of order zero (see, for example, Coker and Filon 1931, p. 319), from the manner of its derivation from 'zero-order' solutions of the biharmonic equation in polar co-ordinates.

Next we consider the solution given by

$$\Omega(z) = D \log z, \quad \omega = \bar{D} z \log z, \quad (11.13)$$

where D is a complex constant. Equations (8.7) and (8.9) show that this gives zero stress resultants round a circuit typical of the connectivity,

whilst (7.4) shows that the solution corresponds to a dislocation of translation $\alpha_1 \mathbf{i} + \alpha_2 \mathbf{j} = \{\alpha_1 + i\alpha_2\} \mathbf{i} = \alpha \mathbf{i}$, given by

$$D = -4\mu i \alpha / \pi(\kappa + 1). \quad (11.14)$$

From the boundary stresses, (11.6) gives

$$4(\widehat{r\bar{r}} - i\bar{r}\widehat{\theta}) = 2D/z,$$

and for the pure dislocation of translation of the annulus, we must annul these stresses over the boundaries $r = a$ and $r = b$. To do this consider the additional complex potentials

$$\Omega(z) = Az^2, \quad \omega(z) = B/z,$$

where A and B are complex constants. The combination gives boundary stresses, using (11.6),

$$4(\widehat{r\bar{r}} - i\bar{r}\widehat{\theta}) = 2z \left\{ \frac{D}{z^2} + \bar{A} - \frac{B}{z^2 z^2} \right\},$$

which vanishes over the boundaries if, for $r = a$ and $r = b$,

$$\frac{D}{r^2} + \bar{A} - \frac{B}{r^4} = 0.$$

Solving for A and B and substituting for D from (11.14), we find that the complex potentials for the pure translational dislocation, often referred to as the pure dislocation of order unity (see Coker and Filon 1931, p. 370) are

$$\left. \begin{aligned} \Omega(z) &= -\frac{4\mu i}{\pi(\kappa + 1)} \left\{ \frac{\bar{\alpha} z^2}{a^2 + b^2} + \alpha \log z \right\}, \\ \omega(z) &= \frac{4\mu i}{\pi(\kappa + 1)} \left\{ \bar{\alpha} z \log z - \alpha \frac{a^2 b^2}{a^2 + b^2} \frac{1}{z} \right\}, \end{aligned} \right\} \quad (11.15)$$

where α is the small complex displacement defining the dislocation of translation.

For the two problems (iv), from (3.6) we put $c = 0$, $d = -\frac{1}{2}n^2$ in equations (11.3)–(11.5), and by adding the last pair and changing the sign of i we find

$$4(\widehat{r\bar{r}} - i\bar{r}\widehat{\theta}) = \Omega'(z) + \bar{\Omega}'(\bar{z}) - z\Omega''(z) - \frac{z}{\bar{z}}\omega''(z) - \rho r n^2 z \bar{z}, \quad (11.16)$$

where $\tau = (\kappa + 3)/(\kappa + 1)$,

and $\tau = (3 - 2\eta)/2(1 - \eta)$ or $(3 + \eta)/2$, (11.17)

according as we are dealing with plane strain or generalized plane stress.

Now take $\Omega(z) = Az$, $\omega(z) = B \log z$,

where A and B are real. With these (11.16) becomes

$$4(\bar{r}\bar{r} - i\bar{r}\bar{\theta}) = 2A + (B/z\bar{z}) - \rho\tau n^2 z\bar{z}. \quad (11.18)$$

Taking $B = 0$, $2A = \rho\tau n^2 a^2$, we have zero stresses on the rim $r = a$ of a complete rotating disk, with finite stresses at the origin, and clearly from (7.4) this is a non-dislocational solution, hence the appropriate complex potentials for the complete uniformly rotating disk or cylinder are

$$\Omega(z) = \frac{1}{2}\rho\tau n^2 a^2 z, \quad \omega(z) = 0. \quad (11.19)$$

Again, choosing A and B to make the right-hand side of (11.18) vanish for $r = a$ and $r = b$, we find that the appropriate complex potentials for the rotating annulus are

$$\Omega(z) = \frac{1}{2}\rho\tau n^2 (a^2 + b^2) z, \quad \omega(z) = -\rho n^2 \tau a^2 b^2 \log z. \quad (11.20)$$

We then find, from (11.18) and (11.4), that

$$\left. \begin{aligned} \bar{r}\bar{r} &= \frac{1}{4}\rho\tau n^2 (a^2 - r^2)(r^2 - b^2)/r^2, \quad \bar{r}\bar{\theta} = 0, \\ \bar{\theta}\bar{\theta} &= \frac{1}{4}\rho\tau \frac{n^2}{r^2} \left\{ a^2 b^2 + (a^2 + b^2)r^2 - \left(3 - \frac{4}{\tau}\right)r^4 \right\}. \end{aligned} \right\} \quad (11.21)$$

This agrees with the well-known solution (e.g. see Timoshenko 1934, p. 68), but disagrees with that given by Bricas (1937, p. 256), which is free of elastic constants. Bricas does not solve for the displacements and discuss their physical admissibility, and although his stresses satisfy the boundary conditions, they correspond to inadmissible (many-valued) displacements.

Solutions for an unstressed hole in an infinite plate

Here we consider:

- (i) an unstressed circular boundary in an infinite plate, with no body forces, and under an all-round tension at infinity,
- (ii) the similar problem for the material at infinity in a state of simple tension making an angle α with the x -axis,
- (iii) similar problems under gravitational body force,
- (iv) similar problems under gravitational body force and with dislocation.

Take the complex potentials

$$\Omega(z) = Az, \quad \omega(z) = B \log z,$$

where A and B are real. Then with no body forces, (11.6) gives

$$4(\bar{r}\bar{r} - i\bar{r}\bar{\theta}) = 2A + B/r^2,$$

so the boundary stresses vanish along $r = a$, if $B = -2a^2A$, and the state of stress at infinity is clearly an all-round tension of amount $T = \frac{1}{2}A$. Accordingly the appropriate complex potentials, since they clearly give no dislocation, from (7.4), are

$$\Omega(z) = 2Tz, \quad \omega(z) = -4a^2T \log z. \quad (11.22)$$

For the second problem take the complex potentials

$$\Omega(z) = Az + B/z, \quad \omega(z) = C \log z + D/z^2 + Ez^2, \quad (11.23)$$

where, of the complex constants, A need only be real, but C must be real, from (8.7) and (8.9), to give zero stress resultants round the boundary. From (7.4) the solution is non-dislocational. Then (11.6) gives

$$4(\bar{r}\bar{r} - i\bar{r}\bar{\theta}) = 2A - 3\frac{B}{z^2} - \frac{\bar{B}}{\bar{z}^2} + \frac{C}{z\bar{z}} - 2E\frac{z}{\bar{z}} - 6\frac{D}{z\bar{z}^3},$$

which becomes, along the boundary $z\bar{z} = a^2$,

$$4(\bar{r}\bar{r} - i\bar{r}\bar{\theta}) = \frac{1}{a^2}(2Aa^2 + C) - \frac{z^2}{a^4}(\bar{B} + 2Ea^2) - \frac{3}{a^2z^2}(Ba^2 + 2D),$$

so that the hole is unstressed if

$$C = -2Aa^2, \quad B = -2\bar{E}a^2, \quad D = \bar{E}a^4. \quad (11.24)$$

Also, from (11.23), (6.14) and (6.15), with no body forces,

$$\bar{x}\bar{x} + \bar{y}\bar{y} = A - \frac{1}{2}\left(\frac{B}{z^2} + \frac{\bar{B}}{\bar{z}^2}\right),$$

$$\bar{x}\bar{x} - \bar{y}\bar{y} + 2i\bar{x}\bar{y} = -\left\{E + \bar{B}\frac{z}{\bar{z}^3} - \frac{1}{2}\frac{C}{z^2} + 3\frac{\bar{D}}{\bar{z}^4}\right\},$$

hence as $r \rightarrow \infty$, the stresses tend to values given by

$$\bar{x}\bar{x} + \bar{y}\bar{y} = A, \quad \bar{x}\bar{x} - \bar{y}\bar{y} + 2i\bar{x}\bar{y} = -\bar{E},$$

and, from (4.9), we have a state of simple tension T at infinity, in a direction making an angle α with the x -axis where,

$$A = T, \quad -\bar{E} = Te^{2i\alpha}. \quad (11.25)$$

Accordingly the appropriate complex potentials for the second problem are, from (11.23)–(11.25),

$$\Omega(z) = T(z + 2a^2e^{2i\alpha}/z), \quad \omega(z) = -T(2a^2 \log z + e^{-2i\alpha}z^2 + a^4e^{2i\alpha}/z^2). \quad (11.26)$$

A dislocational solution for the space outside the circular hole is readily found, but it should be borne in mind that the rotational dislocation will inevitably give rise to infinite displacements at infinity, so that we should restrict the solution to a finite region and imagine the appropriate stresses applied to the boundary of the finite region. Taking the complex potentials

$$\Omega(z) = A \log z + Bz \log z + Cz, \quad \omega(z) = \bar{A}z \log z + D/z,$$

where B and C are real, A and D are complex, we find from (7.4) that the solution corresponds to a dislocation $\alpha + i\beta z$ if

$$A = -4\mu i\alpha/\pi(\kappa + 1), \quad B = 4\mu\beta/\pi(\kappa + 1),$$

whilst from (8.7) and (8.9) the stress resultants round the boundary vanish, there being no body force. Also from (11.6) the boundary stresses are given by

$$4(\bar{r}\bar{r} - i\bar{r}\bar{\theta}) = \frac{2}{z} \left(A - \frac{D}{z\bar{z}} \right) + B \log z\bar{z} + B + 2C,$$

and vanish along $z\bar{z} = a^2$ if

$$D = a^2 A, \quad C = -B(\frac{1}{2} + \log a).$$

Hence the complex potentials for the large plate pierced by a circular hole under no body forces, undergoing the general dislocation $\alpha + i\beta z$, are

$$\left. \begin{aligned} \Omega(z) &= -\frac{4\mu i\alpha}{\pi(\kappa + 1)} \log z + \frac{4\mu\beta}{\pi(\kappa + 1)} \{z \log z - (\frac{1}{2} + \log a)z\}, \\ \omega(z) &= \frac{4\mu i}{\pi(\kappa + 1)} \{\bar{a}z \log z - \alpha a^2/z\}. \end{aligned} \right\} \quad (11.27)$$

When the body force is gravitational, acting along the negative y -axis, we must put $c = -\frac{1}{2}ig$, $d = 0$ (putting $\alpha = -\frac{1}{2}\pi$ in (3.5)), and then, taking complex potentials,

$$\Omega(z) = -\frac{1}{4}i\rho g\gamma z^2, \quad \omega(z) = 0, \quad (11.28)$$

(11.3)–(11.5) give

$$\bar{r}\bar{r} + \bar{\theta}\bar{\theta} = -i\rho g(z - \bar{z}), \quad \bar{r}\bar{r} - \bar{\theta}\bar{\theta} + 2i\bar{r}\bar{\theta} = 0,$$

whence the boundary stresses for the particular solution (11.28) are given by

$$4(\bar{r}\bar{r}_1 - i\bar{r}\bar{\theta}_1) = 2i\rho g(\bar{z} - z). \quad (11.29)$$

From (3.9) other solutions can be formed by adding to this particular solution any solution derived from the equations with no body forces. Suppose we take

$$\Omega_0(z) = iB \log z, \quad \omega_0(z) = iBz \log z,$$

where B is real, and form a complementary solution from the equations with no body forces, the combined solutions give boundary stresses

$$4(\bar{r}\bar{r} - i\bar{r}\bar{\theta}) = 2i(\bar{z} - z)\{\rho g + B/z\bar{z}\},$$

which vanishes over $|z| = a$, if $B = -\rho ga^2$. Hence, substituting for γ in (11.28) from (6.12),

$$\Omega(z) = -i\rho g\left\{a^2 \log z + \frac{\kappa-1}{\kappa+1}z^2\right\}, \quad \omega(z) = -i\rho ga^2 z \log z, \quad (11.30)$$

are complex potentials which make the stresses vanish over the boundary $r = a$. The stresses in the heavy plate pierced by the circular and unstressed hole are then

$$\bar{r}\bar{r} = \rho gr \sin \theta \left(1 - \frac{a^2}{r^2}\right), \quad \bar{\theta}\bar{\theta} = \rho gr \sin \theta, \quad \bar{r}\bar{\theta} = 0. \quad (11.31)$$

These appear to be the results obtained by Bricas (1937, p. 247) in this case. But the result is not satisfactory, since (7.4) identifies this as the solution for a translational dislocation parallel to the x -axis of amount $\pi\rho ga^2(\kappa+1)/4\mu$, which is not apparent in the work of Bricas as he does not in this case discuss the displacements. His solution, identical with (11.31), moreover, purports to be for the body force in the opposite direction to that which we have taken. This discrepancy was eventually traced to incorrect body-stress equations (Bricas, 1937, p. 12; cp. Love 1927, p. 85), in which the body forces have the wrong signs. Quite apart from this error, which would seem to vitiate all his results involving body forces, the general results obtained by Bricas in his chapters XXII and XXIV need a further discussion of the displacements before they can be applied with safety.

To find the correct solution for the heavy infinite plate pierced by a circular hole, with no dislocation and no stresses round the boundary, we take the particular solution (11.28) with the complementary solution

$$\Omega_0(z) = iA \log z, \quad \omega_0(z) = iAkz \log z + iB/z, \quad (11.32)$$

where A and B are real. Then (7.4) shows that this is non-dislocational in character, and (8.7) and (8.9) show that the stress resultants for the boundary $r = a$ in the combined solution are

$$X + iY = i\pi\left\{\frac{1}{2}(\kappa+1)A + \rho ga^2\right\}, \quad M = 0,$$

and so vanish if $A = -2\rho ga^2/(\kappa+1)$, whilst the combined boundary stresses are given, from (11.6), as

$$4(\bar{r}\bar{r} - i\bar{r}\bar{\theta}) = 2i\rho g(\bar{z} - z) + 2i\rho g\left\{\frac{a^2}{\bar{z}} - \frac{2a^2}{\kappa+1}\frac{1}{z}\right\} - \frac{2iB}{z\bar{z}^2}.$$

This vanishes over $z\bar{z} = a^2$ if $B = \rho g a^4 (\kappa - 1)/(\kappa + 1)$, hence the complex potentials for the heavy infinite plate pierced by a circular hole which is unstressed, are, substituting for γ in (11.28) from (6.12),

$$\left. \begin{aligned} \Omega(z) &= -i\rho g \left(\frac{\kappa-1}{\kappa+1} \right) z^2 - 2i\rho g \frac{a^2}{\kappa+1} \log z, \\ \omega(z) &= i\rho g \left(\frac{\kappa-1}{\kappa+1} \right) \frac{a^4}{z} - 2i\rho g \frac{a^2\kappa}{\kappa+1} z \log z. \end{aligned} \right\} \quad (11.33)$$

This corresponds to a state of stress at infinity given by

$$\bar{r}\bar{r} = \rho g y + O(1/r), \quad \bar{r}\bar{\theta} = O(1/r),$$

and we should restrict the solution to a finite region and imagine the appropriate stresses applied to the boundary of this finite region. The complementary solutions (11.22) and (11.28) may be added to (11.33) leaving the hole unstressed but altering the necessary stress system round the other boundary.

The solution to the corresponding problems of type (iv) is obtained simply by combining the dislocational solution (11.27) as a complementary solution to those of type (iii).

Nuclei of strain in an infinite plate

Suppose body force absent and consider the complex potentials

$$\Omega(z) = -\frac{2F}{\pi(\kappa+1)} \log z, \quad \omega(z) = \frac{2F\kappa}{\pi(\kappa+1)} z \log z. \quad (11.34)$$

From (7.4) this is clearly non-dislocational in character throughout an infinite plate, and from (8.7) and (8.9) the stress resultants for a circuit embracing the origin are given by

$$X + iY = -F, \quad M = 0,$$

so that the complex force $X + iY$ applied to the infinite plate is $+F$.

Similarly the complex potentials

$$\Omega(z) = 0, \quad \omega(z) = 2i \frac{G}{\pi} \log z \quad (11.35)$$

give a non-dislocational solution involving stress resultants

$$X + iY = 0, \quad M = -G,$$

so that the stresses apply a couple $+G$ to the infinite plate. Hence (11.34) and (11.35) give the solutions for an isolated force and an isolated couple applied at the origin to an infinite plate under no body force.

Nuclei of strain in a semi-infinite plate

Here we consider the problems of the isolated force and couple applied at the point $z = c$ (c real and positive) in the semi-infinite plate for which x is positive, the boundary $x = 0$ being unstressed and the material under no body force.

Consider first the couple nucleus. Putting $z_1 = z - c$, the complex potentials for the couple G applied at C ($z_1 = 0$) in an infinite plate would be

$$\Omega_C(z_1) = 0, \quad \omega_C(z_1) = 2i \frac{G}{\pi} \log z_1, \quad (11.36)$$

giving rise to boundary stresses which are given, from (6.14) and (6.15), by

$$4(\widehat{xx} + i\widehat{xy}) = \Omega'(z_1) + \bar{\Omega}'(\bar{z}_1) - z_1 \bar{\Omega}''(\bar{z}_1) - \bar{\omega}'(\bar{z}_1) \quad (11.37)$$

or

$$4(\widehat{xx} + i\widehat{xy}) = -2iG/\pi z_1^2.$$

Along the boundary $x = 0$, we can write this

$$4(\widehat{xx} + i\widehat{xy}) = -2iG/\pi z_2^2,$$

where $z_2 = z + c$, since the boundary has the equation $\bar{z}_1 + z_2 = 0$. These stresses have to be annulled by solutions possessing no logarithmic singularities or poles in the positive half-plane. Consider therefore the complex potentials, appropriate to the origin C' , $z_2 = 0$, given by

$$\Omega_{C'}(z_2) = Ai/z_2, \quad \omega_{C'}(z_2) = iB \log z_2 + iCc/z_2, \quad (11.38)$$

where A, B, C are real constants of the same dimensions as G . Since the origin $z_2 = 0$ is a point outside the material these potentials will give admissible stresses and displacements, the boundary stresses being

$$4(\widehat{xx} + i\widehat{xy}) = -\frac{Ai}{z_2^2} + \frac{Ai}{z_2^2} + 2Ai \frac{z_2}{z_2^3} - i \frac{B}{z_2^2} + 2i \frac{Cc}{z_2^3},$$

or, eliminating c from the combined stresses arising from the complex potentials (11.36) and (11.38), since $z_2 + z_1 = 2c$, we find

$$4(\widehat{xx} + i\widehat{xy}) = -i \left(2 \frac{G}{\pi} + A \right) \frac{1}{z_2^2} + i(A - B + C) \frac{1}{z_2^3} + i \frac{z_2}{z_2^3} (2A + C),$$

so that the straight boundary $x = 0$ is free from stress if

$$A = -2 \frac{G}{\pi}, \quad C = -2A = 4 \frac{G}{\pi}, \quad B = A + C = 2 \frac{G}{\pi}.$$

From these values of A, B, C , and using the results of (9.1) to change the origins of the two sets of potentials (11.36) and (11.38) to the origin $O, z = 0$,

the complex potentials appropriate to the couple nucleus of strain in a semi-infinite plate are

$$\Omega(z) = -2i \frac{G}{\pi} \frac{1}{z_2}, \quad \omega(z) = 2i \frac{G}{\pi} \left\{ \log z_1 z_2 + \frac{C}{z_2} \right\}. \quad (11.39)$$

From this and (10.1), Airy's stress function for the problem is the real part of $i \frac{G}{2\pi} \{ \log z_1 z_2 - \bar{z}_1/z_2 \}$, or

$$\chi = -\frac{G}{2\pi} \{ \theta_1 + \theta_2 + 2xy/r_2^2 \}, \quad (11.40)$$

where $z_1 = r_1 e^{i\theta_1}$, $z_2 = r_2 e^{i\theta_2}$. This agrees with the solution due to Ghosh (1937, p. 177).

Next consider the solution for the complex force F applied at the point C , $z = c$. From (11.34) the complex potentials for the infinite plate can be written

$$\Omega_C(z_1) = -E \log z_1, \quad \omega(z_1) = \kappa E z_1 \log z_1, \quad E = \frac{2F}{\pi(\kappa+1)}, \quad (11.41)$$

giving rise to boundary stresses

$$4(\bar{x}\bar{x} + i\bar{x}\bar{y}) = -\left\{ \frac{E}{z_1} + \frac{\bar{E}}{\bar{z}_1} + \frac{z_1}{\bar{z}_1^2} \bar{E} + \frac{\kappa E}{\bar{z}_1} \right\},$$

from (11.37), or, since $z_1 + \bar{z}_2 = 0$ along the boundary $x = 0$,

$$4(\bar{x}\bar{x} + i\bar{x}\bar{y}) = -\left\{ \frac{E}{z_2^2} + (\bar{E} + \kappa E) \frac{1}{z_2} + \bar{E} \frac{\bar{z}_2}{z_2^2} \right\}. \quad (11.42)$$

To annul these boundary stresses, consider the complex potentials, relative to the origin C' , for which $z = -c$, $z_2 = 0$,

$$\Omega_{C'}(z_2) = A \log z_2 + Bc/z_2, \quad \omega_{C'}(z_2) = (Cz_2 + Dc) \log z_2 + Gc^2/z_2, \quad (11.43)$$

which give rise to boundary stresses

$$4(\bar{x}\bar{x} + i\bar{x}\bar{y}) = \frac{A}{z_2} + (\bar{A} - \bar{C}) \frac{1}{\bar{z}_2} - \frac{Bc}{z_2^2} + (\bar{D} - \bar{B}) \frac{c}{\bar{z}_2^2} + \bar{A} \frac{z_2}{\bar{z}_2^2} - 2\bar{B} \frac{cz_2}{\bar{z}_2^2} - 2 \frac{\bar{G}c^2}{\bar{z}_2^2}. \quad (11.44)$$

From the combined solutions derived from the complex potentials (11.41) and (11.43) relative to the two origins C and C' , we find, on eliminating c from the combination of (11.42) and (11.44) by means of $z_2 + \bar{z}_2 = 2c$, that

$$\begin{aligned} 4(\bar{x}\bar{x} + i\bar{x}\bar{y}) = & (\bar{A} - \bar{C} - E - \frac{1}{2}B + \frac{1}{2}\bar{D} - \frac{1}{2}G) \frac{1}{\bar{z}_2} + (A - \bar{E} - \kappa E - \frac{1}{2}B) \frac{1}{z_2} \\ & - (\bar{E} + \frac{1}{2}B) \frac{\bar{z}_2}{z_2^2} + (\bar{A} - \frac{3}{2}\bar{B} + \frac{1}{2}\bar{D} - \bar{G}) \frac{z_2}{\bar{z}_2^2} - (\bar{B} + \frac{1}{2}G) \frac{z_2^2}{\bar{z}_2^2}. \end{aligned}$$

The stresses vanish along the boundary if the expressions in the five brackets vanish. These lead simply to the determination of the five constants A , B , C , D , G as

$$A = \kappa E, \quad B = -2\bar{E}, \quad C = -\bar{E}, \quad D = 2(\bar{E} - \kappa E), \quad G = 4\bar{E}. \quad (11.45)$$

Hence the complex potentials for the problem, relative to the origin O , are from (11.41) and (11.43), using the results of (9.1),

$$\left. \begin{aligned} \Omega(z) &= \frac{2}{\pi(\kappa+1)} \{F \log z_1 + \kappa F \log z_2 - 2\bar{F}c/z_2\}, \\ \omega(z) &= \frac{2}{\pi(\kappa+1)} \{-\kappa\bar{F}z_1 \log z_1 - \bar{F}z_1 \log z_2 - Fc \log z_1 - \kappa F\bar{c} \log z_2 + 2\bar{F}c^2/z_2\}. \end{aligned} \right\} \quad (11.46)$$

From these complex potentials and (10.1), Airy's stress function χ for this problem is the real part of

$$\frac{1}{2\pi(\kappa+1)} \{(F\bar{z}_1 - \kappa\bar{F}z_1) \log z_1 + (\kappa F\bar{z}_1 - \bar{F}z_1) \log z_2 - 2\bar{F}c\bar{z}_1/z_2\}.$$

Since, in generalized plane stress $(1+\kappa)(1+\eta) = 4$, from (6.13), this yields, on dropping an irrelevant constant,

$$\begin{aligned} \chi &= \frac{X}{2\pi} \{y(\theta_1 + \theta_2) - \frac{1}{2}(1-\eta)(x-c) \log(r_1/r_2) - (1+\eta)cx(x+c)/r_2^2\} \\ &+ \frac{Y}{2\pi} \{-(x-c)(\theta_1 + \theta_2) - \frac{1}{2}(1-\eta)y \log(r_1/r_2) + (1+\eta)cxy/r_2^2\}, \end{aligned} \quad (11.47)$$

which agrees with the two results found separately for X and Y by Melan (1932, pp. 345, 346). Comparison of the methods reveals the saving of much labour.

Nuclei of strain in a circular disk

These results, with their general resemblance to hydrodynamical problems possessing the same boundaries and involving the same points of singularity, inevitably suggest the problem of a circular disk, with its boundary free from traction, in equilibrium under the action of two nuclei of strain only, there being no body forces. We shall suppose the disk, of radius a , to be under the action of a force $\mathbf{F} = F\mathbf{i} = Re^{i\alpha}\mathbf{i}$, and a couple $G\mathbf{k}$, both at the centre O , together with a force $-\mathbf{F}$ at C ($z = c$, $c \leq a$; c is real), as in figure 6. For equilibrium of these external forces we must have

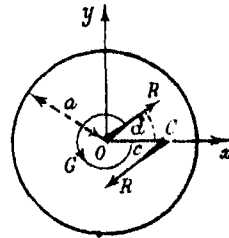


FIGURE 6

$$G = Rc \sin \alpha = -\frac{1}{2}ic(F - \bar{F}),$$

and, if $z_1 = z - c$, the complex potentials for these nuclei of strain are, from (11.34) and (11.35):

$$(i) \text{ for } G, \quad \Omega(z) = 0, \quad \omega(z) = 2i \frac{G}{\pi} \log z = \frac{c}{\pi} (F - \bar{F}) \log z, \quad (11.48)$$

(ii) for F at O ,

$$\Omega(z) = -\frac{2F}{\pi(\kappa+1)} \log z, \quad \omega(z) = \frac{2\bar{F}\kappa}{\pi(\kappa+1)} z \log z, \quad (11.49)$$

(iii) for $-F$ at C , using also (9.1),

$$\left. \begin{aligned} \Omega(z) &= \Omega_C(z_1) = \frac{2F}{\pi(\kappa+1)} \log z_1, \\ \omega(z) &= \omega_C(z_1) - c\Omega(z) = -\frac{2\bar{F}\kappa}{\pi(\kappa+1)} z_1 \log z_1 - \frac{2Fc}{\pi(\kappa+1)} \log z_1. \end{aligned} \right\} \quad (11.50)$$

From (11.6) these lead to stresses $\hat{r}\hat{r}$, $\hat{r}\hat{\theta}$ given by

$$\begin{aligned} 4(\hat{r}\hat{r} - i\hat{r}\hat{\theta}) &= \frac{2F}{\pi(\kappa+1)} \left\{ \frac{1}{z_1} - \frac{2}{z} + \frac{z}{z_1^2} - \frac{cz}{z_1 z^2} \right\} \\ &\quad + \frac{2\bar{F}}{\pi(\kappa+1)} \left\{ \frac{1}{\bar{z}_1} - (\kappa+1) \frac{1}{\bar{z}} + \frac{\kappa z}{z_1 \bar{z}} \right\} + \frac{c}{\pi} (F - \bar{F}) \frac{1}{z\bar{z}}. \end{aligned}$$

Now, along the boundary we have $z\bar{z} = a^2$ and, writing $z_2 = z - a^2/c$, we have

$$z_1 = \frac{a^2}{z} - c = -c \frac{\bar{z}_2}{z},$$

so that

$$\begin{aligned} \frac{1}{z_1} &= -\frac{\bar{z}}{c\bar{z}_2} = -\frac{1}{c} - \frac{a^2}{c^2\bar{z}_2}; \quad \frac{1}{z} = \frac{\bar{z}}{a^2} = \frac{\bar{z}_2}{a^2} + \frac{1}{c}; \\ \frac{z}{z_1^2} &= z\bar{z} \left(\frac{1}{z\bar{z}_1} \right) \frac{1}{z_1} = \frac{a^2}{c\bar{z}_2} \left(\frac{1}{c} + \frac{a^2}{c^2\bar{z}_2} \right) = \frac{a^2}{c^2\bar{z}_2} + \frac{a^4}{c^3\bar{z}_2^2}; \\ \frac{z}{z_1 z^2} &= \frac{1}{z\bar{z}} \left(\frac{z}{z_1} \right)^2 = \frac{1}{z\bar{z}} \left(-\frac{z\bar{z}}{c\bar{z}_2} \right)^2 = \frac{a^2}{c^2\bar{z}_2^2}; \\ \frac{z}{z_1 \bar{z}_1} &= \frac{1}{\bar{z}} \left(-\frac{a^2}{c\bar{z}_2} \right) = \frac{1}{\bar{z}} - \frac{1}{\bar{z}_2} = \frac{z_2}{a^2} + \frac{1}{c} - \frac{1}{\bar{z}_2}; \end{aligned}$$

and with these results the boundary stresses can be interpreted in terms of z_2 , \bar{z}_2 as

$$\begin{aligned} 4(\hat{r}\hat{r} - i\hat{r}\hat{\theta}) &= \frac{2F}{\pi(\kappa+1)} \left\{ \frac{a^2}{c} \left(\frac{a^2}{c^2} - 1 \right) \frac{1}{\bar{z}_2^2} - 2 \frac{\bar{z}_2}{a^2} - \frac{3}{c} + \frac{1}{2} \frac{c}{a^2} (\kappa+1) \right\} \\ &\quad - \frac{2\bar{F}}{\pi(\kappa+1)} \left\{ \frac{\kappa}{\bar{z}_2} + \frac{a^2}{c^2\bar{z}_2} + \frac{z_2}{a^2} + \frac{2}{c} + \frac{1}{2} \frac{c}{a^2} (\kappa+1) \right\}. \quad (11.51) \end{aligned}$$

Consider the complex potentials, relative to the origin O , given by

$$\Omega(z) = A \log z_2 + \frac{B}{z_2} + Cz^2 + Ez, \quad \omega(z) = Gz_2 \log z_2 + K \log z_2 + \frac{L}{z_2}, \quad (11.52)$$

where, of the complex constants $A \dots L$, we may take E to be real. Since the point $z = z_2$ is outside the material, these complex potentials will certainly lead to physically admissible displacements and stresses. We find, using (11.6),

$$4(\bar{r}\bar{r} - ir\bar{\theta}) = \frac{A}{z_2} - \frac{B}{z_2^2} + 2Cz + 2E + 2\bar{C}\bar{z} - \frac{\bar{B}}{\bar{z}_2^2} + \frac{\bar{A}}{\bar{z}_2} - z \left(-\frac{A}{z_2^2} + 2\frac{B}{z_2^3} + 2C \right) - \frac{z}{\bar{z}} \left(\frac{G}{z_2} - \frac{K}{z_2^2} + 2\frac{L}{z_2^3} \right),$$

which, on the boundary $z\bar{z} = a^2$, since $z = z_2 + a^2/c$, can be written in terms of z_2 and \bar{z}_2 as

$$\begin{aligned} 4(\bar{r}\bar{r} - ir\bar{\theta}) &= \frac{\bar{A}}{\bar{z}_2} + \frac{1}{z_2} \left(2A - 2\frac{L}{a^2} + 2\frac{K}{c} - \frac{a^2}{c^2}G \right) + 2\bar{C}\bar{z}_2 - \frac{G}{a^2}z_2 \\ &\quad + \frac{1}{z_2^2} \left(A\frac{a^2}{c} - 3B + \frac{a^2}{c^2}K - 4\frac{L}{c} \right) - \frac{\bar{B}}{\bar{z}_2^2} - \frac{2a^2}{c^2\bar{z}_2^2}(Bc + L) \\ &\quad + \left(2\bar{C}\frac{a^2}{c} + 2E - 2\frac{G}{c} + \frac{K}{a^2} \right). \end{aligned} \quad (11.53)$$

We now add the solutions for the stresses (11.51) and (11.53) arising from the complex potentials (11.48)–(11.50) and (11.52), and derive the values of the constants $A \dots L$ from the condition that the boundary is to be unstressed. From the vanishing of the coefficients of $1/\bar{z}_2$, $1/\bar{z}_2^2$, \bar{z}_2 , z_2 , $1/z_2^2$, we find

$$\left. \begin{aligned} A &= \frac{2F\kappa}{\pi(\kappa+1)}, \quad B = \frac{2\bar{F}}{\pi(\kappa+1)} \frac{a^2}{c^3} (a^2 - c^2), \quad C = \frac{2\bar{F}}{\pi(\kappa+1)} \frac{1}{a^2}, \\ G &= -\frac{2\bar{F}}{\pi(\kappa+1)}, \quad L = -\frac{2\bar{F}}{\pi(\kappa+1)} \frac{a^2}{c^2} (a^2 - c^2), \end{aligned} \right\} \quad (11.54)$$

and the vanishing of the coefficients of $1/z_2$ and $1/z_2^2$ then each lead to

$$K = -\frac{2\bar{F}}{\pi(\kappa+1)} \frac{1}{c} (a^2 - c^2) - \frac{2F\kappa c}{\pi(\kappa+1)}, \quad (11.55)$$

whilst the constant term vanishes if

$$E = \frac{2(F + \bar{F})}{\pi(\kappa+1)} \left(\frac{1}{4} \frac{c}{a^2} (\kappa+1) + \frac{1}{2a^2c} (a^2 - c^2) \right). \quad (11.56)$$

From (11.48)–(11.50) and (11.52), using the results of (11.54)–(11.56), the complex potentials for the problem can be written

$$\Omega(z) = \frac{2F}{\pi(\kappa+1)} \{ \log z_1 - \log z + \kappa \log z_2 + \alpha z \} + \frac{2\bar{F}}{\pi(\kappa+1)} \left\{ \frac{z^2}{a^2} + \frac{a^2}{c^2} (a^2 - c^2) \frac{1}{z_2} + \alpha z \right\}, \quad (11.57)$$

where α is the real constant given by

$$\alpha = \frac{1}{4} \frac{c}{a^2} (\kappa+1) + \frac{1}{2a^2c} (a^2 - c^2), \quad (11.58)$$

and

$$\omega(z) = \frac{2Fc}{\pi(\kappa+1)} \left\{ \frac{1}{2} (\kappa+1) \log z - \kappa \log z_2 - \log z_1 \right\} - \frac{2\bar{F}}{\pi(\kappa+1)} \left\{ \frac{1}{2} (\kappa+1) c \log z - z \log z + \kappa z_1 \log z_1 + z_2 \log z_2 \right. \\ \left. - \frac{1}{c} (a^2 - c^2) \log z_2 + \frac{a^2}{c^2} (a^2 - c^2) \frac{1}{z_2} \right\}. \quad (11.59)$$

The present writer has been unable to find a previous solution of this problem in the literature of the subject.

12. SOLUTIONS IN ELLIPTIC CO-ORDINATES

Here we shall take curvilinear co-ordinates given by

$$z = c \cosh \zeta, \quad \zeta = \xi + i\eta, \quad (12.1)$$

and, assuming no body forces, we consider the problems of

- (i) an unstressed elliptic hole, whose boundary is given by $\xi = \alpha$, in an infinite plate, under an all-round tension at infinity;
- (ii) the similar problem when the state of stress at infinity in the infinite plate is one of a tension making an angle β with the major axis of the elliptic boundary of the hole;
- (iii) a rigid cylinder of elliptic boundary $\xi = \alpha$ filling the hole and turned about $z = 0$ through a small angle β .

From (12.1) we find

$$cd\zeta/dz = \operatorname{cosech} \zeta, \quad zd\zeta/dz = \coth \zeta, \quad (12.2)$$

and substituting these in (6.17) and (6.19), with no body forces, we have

$$4(\xi\bar{\xi} - i\xi\eta) = \Omega'(z) + \bar{\Omega}'(\bar{z}) - \coth \zeta \xi \frac{d\Omega'(z)}{d\zeta} - \frac{1}{c} \operatorname{cosech} \zeta \xi \frac{d\omega'(z)}{d\zeta}. \quad (12.3)$$

In the first problem the boundary conditions are $\xi\bar{\xi} = 0$, $\xi\bar{\eta} = 0$ over $\xi = \alpha$, and the all-round tension T at infinity requires, from (4.6), (6.14) and (6.15), that at infinity

$$\Omega'(z) + \bar{\Omega}'(\bar{z}) \rightarrow 4T, \quad z\bar{\Omega}''(\bar{z}) + \bar{\omega}''(\bar{z}) \rightarrow 0. \quad (12.4)$$

Taking the complex potentials

$$\Omega(z) = Ac \sinh \zeta, \quad \omega(z) = Bc^2 \zeta, \quad (12.5)$$

where A and B are real, we see from (7.4) that the solution is non-dislocational in character, whilst from (8.7) and (8.9) it is clear that the stress resultants round a circuit embracing the hole all vanish. Since $\coth \zeta \rightarrow 1$ and $\operatorname{cosech} \zeta \rightarrow 0$ as $\zeta \rightarrow \infty$, the solution (12.5) satisfies the conditions (12.4) at infinity, if $A = 2T$, and then (12.3) becomes

$$4(\xi\bar{\xi} - i\xi\bar{\eta}) = \operatorname{cosech} \zeta \{2T \coth \zeta \sinh \bar{\zeta} + 2T \cosh \bar{\zeta} \coth^2 \zeta + B \operatorname{cosech} \zeta \coth \zeta\}.$$

This will vanish along the boundary $\bar{\zeta} = 2\alpha - \zeta$ if

$$\operatorname{cosech} \zeta \{2T \cosh 2\alpha + B\} = 0, \quad \text{i.e. if } B = -2T \cosh 2\alpha.$$

Consequently the solution of the first problem is given by the complex potentials

$$\Omega(z) = 2Tc \sinh \zeta, \quad \omega(z) = -2Tc^2 \cosh 2\alpha \zeta. \quad (12.6)$$

This corresponds to an Airy's stress function χ given, from (10.1), by

$$\chi = \frac{1}{4}c^2T(\sinh 2\zeta - 2\zeta \cosh 2\alpha), \quad (12.7)$$

a result due to Inglis (see, for example, Coker and Filon 1931, p. 542).

In the second problem, the stresses $\xi\bar{\xi}$, $\xi\bar{\eta}$ again vanish over $\xi = \alpha$, but from (4.9), (6.14) and (6.15), we see that the conditions at infinity are

$$\Omega'(z) + \bar{\Omega}'(\bar{z}) \rightarrow 2T, \quad z\bar{\Omega}''(\bar{z}) + \omega''(z) \rightarrow -2Te^{-2i\beta}. \quad (12.8)$$

Now take the complex potentials

$$\Omega(z) = Ac \cosh \zeta + Bc \sinh \zeta, \quad (12.9)$$

$$\omega(z) = Cc^2 \zeta + Dc^2 \cosh 2\zeta + Ec^2 \sinh 2\zeta, \quad (12.10)$$

where, since an imaginary A only leads to zero stresses and rigid body displacements, we take A to be real. Again, we take C to be real so that $Cy\omega(z)$ is a pure imaginary for circuits embracing the hole, but B, D, E are complex constants, which we shall write as $B = B_1 + iB_2$, etc., where $B_1, B_2, D_1, D_2, E_1, E_2$ are all real. We then see that these complex potentials give zero stress resultants for circuits about the hole, from (8.7) and (8.9), whilst from (7.4) the corresponding displacements are seen to be non-dislocational in character. Then (12.8) leads to

$$A + B_1 = T, \quad 2(D + E) = -Te^{-2i\beta}. \quad (12.11)$$

Also, from (12.3) the complex potentials (12.9) and (12.10) give boundary stresses

$$4(\xi\bar{\xi} - i\xi\bar{\eta}) = 2A + B \coth \zeta + \bar{B} \coth \bar{\zeta} + B \operatorname{cosech}^2 \zeta \coth \bar{\zeta} \\ + \operatorname{cosech} \bar{\zeta} \{(C + 2E) \operatorname{cosech} \zeta \coth \zeta - 4(D \sinh \zeta + E \cosh \zeta)\},$$

or, on using (12.11) to eliminate A and D ,

$$4(\xi\bar{\xi} - i\xi\bar{\eta}) = \operatorname{cosech} \bar{\zeta} \{(2T - 2B_1 + B \coth \zeta) \sinh \bar{\zeta} + (\bar{B} + B \operatorname{cosech}^2 \zeta) \cosh \bar{\zeta} \\ + (C + 2E) \operatorname{cosech} \zeta \coth \zeta + (4E + 2Te^{-2i\beta}) \sinh \zeta - 4E \cosh \zeta\}.$$

This vanishes along $\xi = \alpha$, i.e. $\bar{\zeta} = 2\alpha - \zeta$, if

$$2 \cosh \zeta \{(T - B_1) \sinh 2\alpha - iB_2 \cosh 2\alpha - 2E\} \\ + 2 \sinh \zeta \{Te^{-2i\beta} - (T - B_1) \cosh 2\alpha + iB_2 \sinh 2\alpha + 2E\} \\ + \operatorname{cosech} \zeta \coth \zeta \{B \cosh 2\alpha + C + 2E\} = 0,$$

which is satisfied if the expressions in the brackets vanish separately. The resulting equations are consistent and demand that

$$(T - B_1) \sinh 2\alpha - 2E_1 = 0, \quad B_2 \cosh 2\alpha + 2E_2 = 0, \\ B_1 \cosh 2\alpha + C + 2E_1 = 0, \\ T \cos 2\beta - (T - B_1) \cosh 2\alpha + 2E_1 = 0, \\ -T \sin 2\beta + B_2 \sinh 2\alpha + 2E_2 = 0,$$

hence, solving for B_1 , B_2 , E_1 , E_2 , C , we readily find

$$C = -T(\cosh 2\alpha - \cos 2\beta),$$

$$B = T(1 - e^{2\alpha + 2i\beta}), \quad E = \frac{1}{2}Te^{2\alpha} \sinh 2(\alpha + i\beta),$$

and then from (12.11),

$$A = Te^{2\alpha} \cos 2\beta, \quad D = -\frac{1}{2}Te^{2\alpha} \cosh 2(\alpha + i\beta).$$

Accordingly the solutions for the complex potentials for the unstressed elliptic hole in an infinite plate under a tension at infinity in a direction making an angle β with the major axis of the hole, are

$$\Omega(z) = Tce^{2\alpha} \cos 2\beta \cosh \zeta + Tc(1 - e^{2\alpha + 2i\beta}) \sinh \zeta, \quad (12.12)$$

$$\omega(z) = -Tc^2(\cosh 2\alpha - \cos 2\beta)\zeta - \frac{1}{2}Tc^2e^{2\alpha} \cosh 2(\zeta - \alpha - i\beta). \quad (12.13)$$

The relative simplicity of these results is in strong contrast to the apparent complexity of the Airy stress function, built up by similar tentative methods,

but not given explicitly by Coker and Filon (1931, pp. 542-4, 548-51). As some check, we deduce the peripheral stress $\bar{\eta}\eta_{\xi=\alpha}$. We have

$$\xi\bar{\xi} + \bar{\eta}\eta = \text{real part of } \Omega'(z),$$

whence $\bar{\eta}\eta_{\xi=\alpha} = \text{real part of } T\{e^{2\alpha} \cos 2\beta + (1 - e^{2\alpha+2i\beta}) \coth(\alpha + i\eta)\}$,

$$\text{or } \bar{\eta}\eta_{\xi=\alpha} = T\{\sinh 2\alpha + \cos 2\beta - e^{2\alpha} \cos 2(\beta - \eta)\} / (\cosh 2\alpha - \cos 2\eta), \quad (12.14)$$

agreeing with Coker and Filon (1931, p. 550).

In the third problem, we have an example of boundary conditions in terms of the displacements instead of the stresses. At infinity the displacements must not be infinite, whilst over the elliptic hole $\xi = \alpha$, the displacements are given by

$$u + iv = i\beta z = i\beta c \cosh \zeta,$$

so that over the boundary of the elliptic hole where $\zeta = 2\alpha - \bar{\zeta}$, we must have

$$8\mu(u + iv) = 8\mu i\beta c \{\cosh 2\alpha \cosh \bar{\zeta} - \sinh 2\alpha \sinh \bar{\zeta}\}. \quad (12.15)$$

Consider the complex potentials

$$\Omega(z) = Ae^{-\zeta}, \quad \omega(z) = iBc\zeta + Cce^{-2\zeta}, \quad (12.16)$$

where B is real. From (7.4) this solution is non-dislocational in character, and from (8.7) and (8.9) the stress resultants for any circuit embracing the ellipse $\xi = \alpha$ are

$$X = Y = 0, \quad M = -\frac{\pi}{2} Bc. \quad (12.17)$$

From (6.11) the displacements are given by

$$8\mu(u + iv) = \kappa Ae^{-\zeta} + \bar{A}e^{-\bar{\zeta}} \operatorname{cosech} \bar{\zeta} \cosh \zeta + (iB + 2\bar{C}e^{-2\bar{\zeta}}) \operatorname{cosech} \bar{\zeta},$$

which satisfactorily tend to zero at infinity, where ξ becomes infinite. On the boundary $\zeta = 2\alpha - \bar{\zeta}$, this can be written

$$\begin{aligned} 8\mu(u + iv) &= \kappa Ae^{-2\alpha} + \bar{A}e^{-\bar{\zeta}} \operatorname{cosech} \bar{\zeta} \cosh (2\alpha - \bar{\zeta}) + (iB + 2\bar{C}e^{-2\bar{\zeta}}) \operatorname{cosech} \bar{\zeta} \\ &= \cosh \bar{\zeta} \{\kappa Ae^{-2\alpha} - \bar{A}e^{2\alpha} - 4\bar{C}\} + \sinh \bar{\zeta} \{\kappa Ae^{-2\alpha} + \bar{A}e^{2\alpha} + 4\bar{C}\} \\ &\quad + \operatorname{cosech} \bar{\zeta} \{iB + \bar{A} \cosh 2\alpha + 2\bar{C}\}. \end{aligned}$$

and this can be made to satisfy the boundary condition (12.15) if we choose

$$\begin{aligned} \kappa Ae^{-2\alpha} - \bar{A}e^{2\alpha} - 4\bar{C} &= 8\mu i\beta c \cosh 2\alpha, \\ \kappa Ae^{-2\alpha} + \bar{A}e^{2\alpha} + 4\bar{C} &= -8\mu i\beta c \sinh 2\alpha, \\ iB + \bar{A} \cosh 2\alpha + 2\bar{C} &= 0, \end{aligned}$$

the solution of these giving

$$A = 4\mu i\beta c/\kappa, \quad B = 2\mu\beta c\{e^{2\alpha} + e^{-2\alpha}/\kappa\}, \quad C = \mu i\beta c(\kappa - 1)e^{2\alpha}/\kappa.$$

Hence the appropriate complex potentials for this problem are

$$\Omega(z) = 4\frac{\mu}{\kappa}i\beta c e^{-\zeta}, \quad \omega(z) = 2\frac{\mu}{\kappa}i\beta c^2\{\kappa e^{2\alpha} + e^{-2\alpha}\}\zeta + \frac{\mu}{\kappa}i\beta c^2(\kappa - 1)e^{2\alpha - 2\zeta}, \quad (12.18)$$

and the couple which must be supplied by the rigid ellipse filling the hole $\xi = \alpha$ to the rest of the material is $-M$ or $\mu\beta c^2\{e^{2\alpha} + e^{-2\alpha}/\kappa\}$.

The solution can be identified with that given by Love (1927, p. 273) or Edwardes (1893, p. 270).

13. SOLUTIONS IN BIPOLAR CO-ORDINATES

Here the co-ordinates are given by

$$z = c \tan \frac{1}{2}\zeta, \quad \zeta = \xi + i\eta, \quad (13.1)$$

and with these co-ordinates we investigate solutions, assuming no body force, for

- (i) the ring space between two eccentric circles under internal pressure p_i and external pressure p_o ,
- (ii) the eccentric ring space under uniform boundary shears,
- (iii) the pure dislocational solutions for the ring space,
- (iv) a circular region under equal and opposite forces at the ends of a chord, the circular boundary being otherwise unstressed,
- (v) an infinite region containing an unstressed circular hole, save for equal and opposite forces at the ends of a chord.

Of these problems, the solution for (ii) appears to be new, that for (v) has been given for the particular case when the chord is a diameter.

$$\text{If} \quad z - ic = z_1 = r_1 e^{i\theta_1}, \quad z + ic = z_2 = r_2 e^{i\theta_2}, \quad (13.2)$$

then (13.1) is equivalent to

$$z_1/z_2 = e^{i(\xi - \pi)}, \quad \text{or} \quad \xi - \pi = -i \log(z_1/z_2), \quad (13.3)$$

whence

$$\xi - \pi + i\eta = \theta_1 - \theta_2 - i \log(r_1/r_2),$$

so that the curves $\xi = \text{constant}$ are circles with $z_1 = 0$, $z_2 = 0$ as common points, and the orthogonal family $\eta = \text{constant}$ are circles with these points as real limiting points, given by $\eta = +\infty$, and $\eta = -\infty$ respectively. We

take ξ positive to correspond to x positive, the segment of the y -axis between the limiting points to be given by $\xi = 0$, then outside this segment, $\xi = \pm \pi$ along the y -axis, so that there is a discontinuity of 2π in ξ on crossing the y -axis outside the segment connecting the limiting points. For η positive the circles $\eta = \text{constant}$ surround the limiting point $z = ic$, for η negative they surround the point $z = -ic$ (see figure 7).

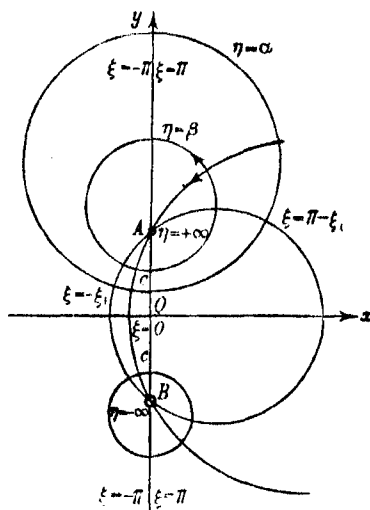


FIGURE 7

The eccentric annulus is defined by $\alpha \leq \eta \leq \beta$, and is the ring space between circles of radii a, b with their centres a distance f apart, where

$$a = c \operatorname{cosech} \alpha, \quad b = c \operatorname{cosech} \beta, \quad f = c (\coth \alpha - \coth \beta). \quad (13.4)$$

From (13.1) $cd\zeta/dz = 1 + \cos \zeta, \quad zd\zeta/dz = \sin \zeta,$ (13.5)

and so the equations (6.17) and (6.19) become, when there are no body forces,

$$\xi\bar{\xi} + \eta\bar{\eta} = \frac{1}{2} \{ \Omega'(z) + \bar{\Omega}'(\bar{z}) \}, \quad (13.6)$$

$$\xi\bar{\xi} - \eta\bar{\eta} + 2i\xi\bar{\eta} = -\frac{1}{2} \left\{ \sin \zeta \frac{d\bar{\Omega}'(\bar{z})}{d\bar{\zeta}} + (1 + \cos \zeta) \frac{1}{c} \frac{d\bar{\omega}'(\bar{z})}{d\bar{\zeta}} \right\}, \quad (13.7)$$

The boundary stresses for the annulus are $\eta\bar{\eta}$ and $\xi\bar{\xi}$, and from (13.6) and (13.7) they can be combined as

$$4(\eta\bar{\eta} + i\xi\bar{\eta}) = \Omega'(z) + \bar{\Omega}'(\bar{z}) + \sin \zeta \frac{d\Omega'(z)}{d\zeta} + (1 + \cos \zeta) \frac{1}{c} \frac{d\omega'(z)}{d\zeta}. \quad (13.8)$$

Consider the complex potentials

$$\Omega(z) = -2pz, \quad \omega(z) = 0 \quad (13.9)$$

where p is real. Equation (13.8) gives

$$\widehat{\eta\eta} + i\widehat{\xi\eta} = -p, \quad (13.10)$$

so that (13.9) gives the solution for annulus under the same internal and external pressure p .

Consider next the complex potentials

$$\Omega(z) = 0, \quad \omega(z) = 2Cic^2\zeta, \quad (13.11)$$

where C is real. From (13.3), (7.4), (8.7) and (8.9) we see that these give a non-dislocational solution with zero stress resultants for closed circuits reducible to the boundaries of the annulus. Equation (13.8) then gives boundary stresses.

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = -2iC \sin \zeta (1 + \cos \bar{\zeta}) = C\{\sinh 2\eta - i(\sin 2\xi + 2 \sin \zeta)\}. \quad (13.12)$$

We now seek a combination of complex potentials which will make the boundary stresses constant. Consider the functions

$$S(z) = c \sin \zeta, \quad C(z) = c \cos \zeta, \quad (13.13)$$

which make

$$S'(z) = \frac{1}{2}(1 + \cos \zeta + \cos 2\zeta), \quad C'(z) = -\frac{1}{2}(2 \sin \zeta + \sin 2\zeta), \quad (13.14)$$

$$\frac{dS'}{d\zeta} = -(\sin \zeta + \sin 2\zeta), \quad \frac{dC'}{d\zeta} = -(\cos \zeta + \cos 2\zeta). \quad (13.15)$$

Hence, if we take the complex potentials

$$\Omega(z) = Ac \sin \zeta, \quad \omega(z) = Ac^2 \cos \zeta, \quad (13.16)$$

where A is real, they clearly give rise to non-dislocational solutions involving zero stress resultants round a closed circuit typical of the connectivity of the eccentric ring space, and (13.8) leads to boundary stresses, making use of (13.13)–(13.16),

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = A\{1 - \cos(\zeta - \bar{\zeta}) + \frac{1}{2}(\cos 2\bar{\zeta} - \cos 2\zeta) + 2 \sin \zeta \sin(\zeta - \bar{\zeta})\}$$

$$\text{or } 4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = A\{1 - \cosh 2\eta + i \sinh 2\eta(\sin 2\xi + 2 \sin \zeta)\}, \quad (13.17)$$

in which the expression in the inner bracket is the same combination which occurs in the inner bracket in (13.12). The same combination also occurs in the stresses derived from the complex potentials

$$\Omega(z) = iBc \cos \zeta, \quad \omega(z) = -iBc^2 \sin \zeta, \quad (13.18)$$

where B is real, since from (13.8), using (13.13)–(13.15), we find

$$4(\eta\eta + i\xi\eta) = iB\{\sin(\zeta - \bar{\zeta}) + \frac{1}{2}(\sin 2\zeta + \sin 2\bar{\zeta}) + 2\sin \zeta \cos(\zeta - \bar{\zeta})\}$$

or $4(\eta\eta + i\xi\eta) = B\{-\sinh 2\eta + i \cosh 2\eta (\sin 2\xi + 2\sin \zeta)\}. \quad (13.19)$

If we now combine the solutions given by the complex potentials (13.11), (13.16) and (13.18), the boundary stresses are given, from (13.12), (13.17) and (13.19), by

$$4(\eta\eta + i\xi\eta) = A(1 - \cosh 2\eta) - (B - C) \sinh 2\eta \\ + i(\sin 2\xi + 2\sin \zeta) \{A \sinh 2\eta + B \cosh 2\eta - C\},$$

which makes

$$\xi\eta = 0, \quad \eta\eta = -\frac{1}{4}\{A(\cosh 2\eta - 1) + (B - C) \sinh 2\eta\} = -P(\eta), \quad (13.20)$$

over $\eta = \alpha$, $\eta = \beta$, if $A \sinh 2\eta + B \cosh 2\eta - C = 0$

for $\eta = \alpha$, $\eta = \beta$, whence

$$A = -C \frac{\sinh(\alpha + \beta)}{\cosh(\beta - \alpha)}, \quad B = C \frac{\cosh(\alpha + \beta)}{\cosh(\beta - \alpha)}, \quad (13.21)$$

leading to

$$P(\alpha) = \frac{1}{2}C \sinh^2 \alpha \tanh(\beta - \alpha), \quad P(\beta) = -\frac{1}{2}C \sinh^2 \beta \tanh(\beta - \alpha). \quad (13.22)$$

The complex potentials of the combined solution are

$$\Omega(z) = icC \frac{\cos[\zeta - i(\alpha + \beta)]}{\cosh(\beta - \alpha)}, \quad \omega(z) = 2Cic^2\zeta - ic^2C \frac{\sin[\zeta - i(\alpha + \beta)]}{\cosh(\beta - \alpha)} \quad (13.23)$$

and the solution is that for the eccentric ring space under an internal pressure $P(\beta)$ and an external pressure $P(\alpha)$. To obtain the solution for the ring space under any internal pressure p_i and external pressure p_o we need only combine the solutions given by (13.9) and (13.23), since (13.10) and (13.20)–(13.22) make

$$p_i = p + P(\beta), \quad p_o = p + P(\alpha),$$

whence
$$p = \frac{p_i \sinh^2 \alpha + p_o \sinh^2 \beta}{\sinh^2 \alpha + \sinh^2 \beta}, \quad C = \frac{2(p_o - p_i) \coth(\beta - \alpha)}{\sinh^2 \alpha + \sinh^2 \beta} \quad (13.24)$$

which can be expressed in terms of a, b, f from (13.4). The complex potentials for the eccentric annulus under internal pressure p_i and external pressure p_o are accordingly

$$\Omega(z) = -\frac{2}{\sinh^2 \alpha + \sinh^2 \beta} \times \left\{ (p_i \sinh^2 \alpha + p_o \sinh^2 \beta) z - ic \frac{(p_o - p_i)}{\sinh(\beta - \alpha)} \cos[\zeta - i(\alpha + \beta)] \right\}, \quad (13.25)$$

$$\omega(z) = \frac{2(p_o - p_i) \coth(\beta - \alpha)}{\sinh^2 \alpha + \sinh^2 \beta} \left\{ 2ic^2 \zeta - ic^2 \frac{\sin[\zeta - i(\alpha + \beta)]}{\cosh(\beta - \alpha)} \right\}. \quad (13.26)$$

This problem was first solved by Jeffery (1921, p. 283), and is given in Coker and Filon (1931, p. 311). The Airy stress function is not there given explicitly, and the easier way to compare results is to find the peripheral stress $\xi\bar{\xi}$. We have, using (13.6), (13.9) and (13.23),

$$\begin{aligned} \xi\bar{\xi} + \eta\bar{\eta} &= \text{real part of } -2p - iC \sin[\zeta - i(\alpha + \beta)] (1 + \cos \zeta) / \cosh(\beta - \alpha) \\ &= -2p + \frac{C}{\cosh(\beta - \alpha)} \{ \cos \xi (1 + \cos \xi \cosh \eta) \sinh(\eta - \alpha - \beta) \\ &\quad - \sin^2 \xi \sinh \eta \cosh(\eta - \alpha - \beta) \}, \end{aligned}$$

so that over $\eta = \alpha$, for example, using (13.24),

$$\begin{aligned} \xi\bar{\xi}_{\eta=\alpha} &= p_o - \frac{2(p_i \sinh^2 \alpha + p_o \sinh^2 \beta)}{\sinh^2 \alpha + \sinh^2 \beta} \\ &\quad - \frac{2(p_o - p_i)}{\sinh^2 \alpha + \sinh^2 \beta} \{ \cos^2 \xi + (\cos \xi \sinh \beta + \sinh \alpha \cosh \beta) \operatorname{cosech}(\beta - \alpha) \} \end{aligned}$$

or

$$\begin{aligned} \xi\bar{\xi}_{\eta=\alpha} &= -p_o - \frac{2(p_o - p_i)}{\sinh^2 \alpha + \sinh^2 \beta} \{ \cos^2 \xi + \cos \xi \sinh \beta \operatorname{cosech}(\beta - \alpha) \\ &\quad + \sinh \alpha \cosh \alpha \coth(\beta - \alpha) \} \\ &= -p_o - \frac{2(p_o - p_i)}{\sinh^2 \alpha + \sinh^2 \beta} (\cosh \alpha + \cos \xi) \{ \sinh \alpha \coth(\beta - \alpha) + \cos \xi \}, \end{aligned} \quad (13.27)$$

which agrees with Jeffery's result for the peripheral stress, when due account is taken of differences in the co-ordinates used.

Consider next the solution for the eccentric annulus under boundary shears uniformly applied so that the annulus is in equilibrium under equal and opposite couples M . From (7.4) the solution given by the complex potentials

$$\Omega(z) = 0, \quad \omega(z) = 2iCc^2 \log(z_1 z_2 / 2c^2) = -2iCc^2 \log(1 + \cos \zeta), \quad (13.28)$$

is non-dislocational in character, and from (8.7) and (8.9) the stress resultants round a circuit reducible to the inner boundary are

$$X = Y = 0, \quad M = -\pi Cc^2. \quad (13.29)$$

The boundary stresses are then given, from (13.8) and (13.28), as

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = 2iC \cos \zeta (1 + \cos \bar{\zeta}) = iC\{\cosh 2\eta + (\cos 2\xi + 2 \cos \zeta)\}. \quad (13.30)$$

We now seek complex potentials which make $\widehat{\eta\eta}$ vanish and $\widehat{\xi\eta}$ constant along the boundaries $\eta = \alpha$, $\eta = \beta$, when taken in combination with (13.28), and give physically admissible stresses and displacements. Consider the complex potentials

$$\Omega(z) = -Ac \cos \zeta, \quad \omega(z) = Ac^2 \sin \zeta, \quad (13.31)$$

where A is real. Using (13.13)–(13.15), (13.8) gives boundary stresses

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = -A\{\sin(\zeta - \bar{\zeta}) + \frac{1}{2}(\sin 2\zeta - \sin 2\bar{\zeta}) + 2 \cos \zeta \sin(\zeta - \bar{\zeta})\},$$

$$\text{or } 4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = -iA\{\sinh 2\eta + \sinh 2\eta(\cos 2\xi + 2 \cos \zeta)\}. \quad (13.32)$$

Again, the complex potentials

$$\Omega(z) = Bci \sin \zeta, \quad \omega(z) = Bc^2i \cos \zeta, \quad (13.33)$$

where B is real, give, from (13.13)–(13.15), using (13.8)

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = -iB\{\cos(\zeta - \bar{\zeta}) + \frac{1}{2}(\cos 2\zeta + \cos 2\bar{\zeta}) + 2 \cos \zeta \cos(\zeta - \bar{\zeta})\},$$

$$\text{or } 4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = -iB\{\cosh 2\eta + \cosh 2\eta(\cos 2\xi + 2 \cos \zeta)\}. \quad (13.34)$$

The combination of the complex potentials (13.28), (13.31) and (13.33) yields boundary stresses, using (13.30), (13.32) and (13.34), given by

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = -i\{A \sinh 2\eta + (B - C) \cosh 2\eta \\ + (A \sinh 2\eta + B \cosh 2\eta - C) (\cos 2\xi + \cos 2\zeta)\}$$

which becomes

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = -i\{A \sinh 2\eta + (B - C) \cosh 2\eta\}, \quad (13.35)$$

along $\eta = \alpha$, $\eta = \beta$, if A and B are given in terms of C , α and β by (13.21). Hence, using (13.29) and (13.4), we have, along the boundary $\eta = \alpha$,

$$\left. \begin{aligned} \widehat{\eta\eta} &= 0, \quad \widehat{\xi\eta} = -\frac{M}{2\pi c^2} \sinh^2 \alpha = -\frac{M}{2\pi a^2}, \\ \text{and } \widehat{\eta\eta} &= 0, \quad \widehat{\xi\eta} = -\frac{M}{2\pi c^2} \sinh^2 \beta = -\frac{M}{2\pi b^2}, \end{aligned} \right\} \quad (13.36)$$

along the boundary $\eta = \beta$.

The complex potentials for the solution of this problem are therefore

$$\Omega(z) = -iM \sin [\zeta - i(\alpha + \beta)] / \pi c \cosh (\beta - \alpha), \quad (13.37)$$

$$\omega(z) = 2i \frac{M}{\pi} \log (1 + \cos \zeta) - iM \cos [\zeta - i(\alpha + \beta)] / \pi \cosh (\beta - \alpha). \quad (13.38)$$

The present writer has been unable to find a previous solution of this problem in the literature of the subject.

We consider next the dislocational solutions for the annulus. Taking the complex potentials

$$\Omega(z) = iA\zeta z, \quad \omega(z) = iAc^2\zeta, \quad (13.39)$$

where A is a real constant, we see from (8.7) and (8.9) that this solution gives zero stress resultants round the boundaries, but that, from (7.4), it corresponds to a rotational dislocation of small angle γ given by

$$A = 4\mu\gamma/\pi(\kappa + 1). \quad (13.40)$$

From (13.8) the boundary stresses are given by

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = iA\{\zeta - \bar{\zeta} - \sin(\zeta - \bar{\zeta})\},$$

$$\text{whence} \quad \widehat{\xi\eta} = 0, \quad \widehat{\eta\eta} = \frac{1}{4}A\{\sinh 2\eta - 2\eta\}, \quad (13.41)$$

so that the solution corresponds to *tensions* p_i , p_o round the boundaries $\eta = \beta$, $\eta = \alpha$ respectively, given by

$$p_i = \frac{1}{4}A\{\sinh 2\beta - 2\beta\}, \quad p_o = \frac{1}{4}A\{\sinh 2\alpha - 2\alpha\}.$$

If this solution is combined with the solution for the annulus under corresponding uniform *pressures* p_i , p_o , the complex potentials for which are given by (13.25), (13.26), we obtain the complex potentials for the pure rotational dislocation of the annulus as

$$\Omega(z) = iA\zeta z - 2Pz + icC \cos [\zeta - i(\alpha + \beta)] \operatorname{sech} (\beta - \alpha), \quad (13.42)$$

$$\omega(z) = iAc^2\zeta + 2iCc^2\zeta - ic^2C \sin [\zeta - i(\alpha + \beta)] \operatorname{sech} (\beta - \alpha), \quad (13.43)$$

$$\text{where} \quad P = \frac{1}{4}A \frac{\sinh \alpha \sinh \beta \sinh (\alpha + \beta) - (\beta \sinh^2 \alpha + \alpha \sinh^2 \beta)}{\sinh^2 \alpha + \sinh^2 \beta}, \quad (13.44)$$

$$C = A \frac{(\beta - \alpha) \coth (\beta - \alpha) - \cosh (\alpha + \beta) \cosh (\beta - \alpha)}{\sinh^2 \alpha + \sinh^2 \beta}, \quad (13.45)$$

and A is given by (13.40).

Consider next the complex potentials

$$\Omega(z) = Bc\zeta, \quad \omega(z) = -Bc\zeta z, \quad (13.46)$$

where B is a real constant. From (8.7) and (8.9) these give zero stress resultants round the boundaries, but, from (7.4), there is an x -resolute of translational dislocation of amount u_0 , given by

$$B = 4\mu u_0 / \pi c(\kappa + 1). \quad (13.47)$$

From (13.8), the boundary stresses are

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = B\{1 - \cos(\zeta - \bar{\xi})\}$$

or

$$\widehat{\xi\eta} = 0, \quad \widehat{\eta\eta} = -\frac{1}{2}B \sinh^2 \eta.$$

Hence the complex potentials (13.46) with B given by (13.47) give a dislocational solution corresponding to tensions p_i , p_o round the boundaries $\eta = \beta$, $\eta = \alpha$ respectively, given by

$$p_i = -\frac{1}{2}B \sinh^2 \beta, \quad p_o = -\frac{1}{2}B \sinh^2 \alpha.$$

A combination of this solution with that given by (13.9) is given by Bricas (1937, p. 216) as the solution of Jeffery's problem (13.25), (13.26); the dislocational character of the solution appears to have been overlooked because the displacements have not been investigated.

If the solution (13.46) is combined with the solution (13.25), (13.26) for the annulus under uniform pressures p_i , p_o , we have the appropriate complex potentials for the x -resolute of pure translational dislocation for the annulus, namely

$$\Omega(z) = Bc\zeta - 2Pz + iCc \cos[\zeta - i(\alpha + \beta)] \operatorname{sech}(\beta - \alpha), \quad (13.48)$$

$$\omega(z) = -Bc\zeta z + 2iCc^2\zeta - ic^2C \sin[\zeta - i(\alpha + \beta)] \operatorname{sech}(\beta - \alpha), \quad (13.49)$$

where
$$P = -\frac{4\mu u_0}{\pi c(\kappa + 1)} \frac{\sinh^2 \alpha \sinh^2 \beta}{\sinh^2 \alpha + \sinh^2 \beta}, \quad (13.50)$$

$$C = \frac{4\mu u_0}{\pi c(\kappa + 1)} \frac{\sinh(\beta + \alpha) \cosh(\beta - \alpha)}{\sinh^2 \alpha + \sinh^2 \beta}, \quad (13.51)$$

and B is given by (13.47).

Next we consider the complex potentials

$$\Omega(z) = iCc\zeta, \quad \omega(z) = iCc\zeta z, \quad (13.52)$$

which give zero stress resultants round the boundaries, from (8.7) and (8.9), whilst from (7.4), they correspond to a y -resolute of translational dislocation for the annulus, of amount v_0 , given by

$$C = 4\mu v_0 / \pi c(\kappa + 1), \quad (13.53)$$

with boundary stresses, from (13.8), given by

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = iC\{1 + 2\cos\zeta + \cos(\zeta + \bar{\zeta})\}. \quad (13.54)$$

Consider also the complex potentials

$$\Omega(z) = 0, \quad \omega(z) = iDc^2 \log(1 + \cos\zeta). \quad (13.55)$$

From (7.4) this is a non-dislocational solution, from (8.7) it corresponds to zero force resultant round the boundary, but, from (8.9), we have a couple resultant M given by

$$D = 2M/\pi c^2. \quad (13.56)$$

This corresponds to boundary stresses

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = -iD\{\cos\zeta + \cos\zeta \cos\bar{\zeta}\}, \quad (13.57)$$

from (13.8). If now we put $D = 2C$ and combine the complex potentials (13.52) and (13.55), so that

$$\Omega(z) = iCc\zeta, \quad \omega(z) = iCc\zeta z + 2iCc^2 \log(1 + \cos\zeta), \quad (13.58)$$

and (13.54) and (13.57) combine to give boundary stresses

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = iC\{1 - \cos(\zeta - \bar{\zeta})\},$$

whence

$$\widehat{\eta\eta} = 0, \quad \widehat{\xi\eta} = -\frac{1}{2}C \sinh^2 \eta, \quad (13.59)$$

and the solution corresponds to a state of strain involving the y -resolute v_0 of dislocation of the annulus, maintained by opposing couples of torque

$$M = 4\mu cv_0/(\kappa + 1), \quad (13.60)$$

arising from uniformly applied boundary shears. If now we combine the solution (13.58) with that given by (13.37) and (13.38), in which M is replaced by $-4\mu cv_0/(\kappa + 1)$, we find

$$\Omega(z) = \frac{4\mu iv_0}{\pi(\kappa + 1)} \{\zeta - \sin[\zeta - i(\alpha + \beta)] \operatorname{sech}(\beta - \alpha)\}, \quad (13.61)$$

$$\omega(z) = \frac{4\mu iv_0 c}{\pi(\kappa + 1)} \{\zeta \tan \frac{1}{2}\zeta - \cos[\zeta - i(\alpha + \beta)] \operatorname{sech}(\beta - \alpha) + 4 \log(1 + \cos\zeta)\}, \quad (13.62)$$

as the complex potentials giving the y -resolute of the pure translational dislocation of the annulus. The Airy stress functions for the three components of the pure dislocation of the annulus were obtained by Ghosh (1926, p. 185).

The fourth problem proposed in this section is an example of a different type, for a circular region defined by the arcs $\xi = -\alpha$, $\xi = \pi - \alpha$, under equal and opposite forces applied at the ends of the chord joining the limiting points $z = \pm ic$. By suitably varying c and α , we have the solution of the problem for any chord of any circle. The boundary stresses for this region are, in the absence of body force,

$$4(\xi\bar{\xi} - i\xi\bar{\eta}) = \Omega'(z) + \bar{\Omega}'(\bar{z}) - \sin \bar{\xi} \frac{d\Omega'(z)}{d\xi} - (1 + \cos \bar{\xi}) \frac{1}{c} \frac{d\omega'(z)}{d\xi}. \quad (13.63)$$

Consider the solution given by the complex potentials

$$\Omega(z) = Ac\xi - Bz, \quad \omega(z) = Ac\xi z, \quad (13.64)$$

where A and B are real. From (13.63)

$$4(\xi\bar{\xi} - i\xi\bar{\eta}) = A - 2B - A \cos(\xi + \bar{\xi}),$$

which vanishes along the boundary $\xi + \bar{\xi} = -2\alpha$ or $2\pi - 2\alpha$, if

$$B = A \sin^2 \alpha. \quad (13.65)$$

The complex potentials possess no cyclic functions for circuits within the material, but have singularities at $z = \pm ic$, corresponding to isolated forces at these points on the boundary. Consider the stress resultants round the limiting semicircles about these points and within the material. These are clearly one-half of the values obtained from the complete limiting circles. Hence from (8.7) and (8.9), we find at $A(0, ic)$ (see figure 8)

$$X - iY = \frac{1}{2}i\pi Ac, \quad M = 0,$$

i.e. we have a force F acting upon the disk at A along the positive y -axis, given by

$$F = -Y = \frac{1}{2}\pi Ac,$$

whilst at $B(0, -ic)$, since description of the limiting semicircle in the same clockwise sense now makes ξ decrease, we have

$$X - iY = -\frac{1}{2}i\pi Ac, \quad M = 0,$$

and hence at B we have an equal and opposite force F along the negative y -axis. The appropriate complex potentials for the circular disk under equal and opposite forces F at the ends of a chord are therefore

$$\Omega(z) = \frac{2F}{\pi} \left\{ \xi - \sin^2 \alpha \tan \frac{1}{2} \xi \right\}, \quad \omega(z) = \frac{2F}{\pi} c \xi \tan \frac{1}{2} \xi. \quad (13.66)$$

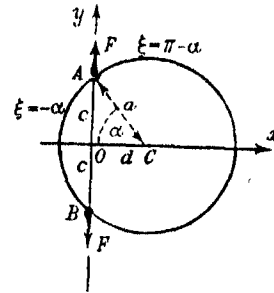


FIGURE 8

The problem was first solved by Hertz (1883), for the case where the forces are at the ends of a diameter, later by Michell (1900), and also by Bricas (1937, p. 227), who employs curvilinear co-ordinates similar to those used here.

Let us examine the nature of the solution (13.66) when applied to the space outside the circle $\xi = -\alpha$, $\pi - \alpha$, now taken as the boundary of a circular hole in an infinite plate. There will be localized forces $-Fj$ at $z = ic$, Fj at $z = -ic$, otherwise the hole is unstressed.

From (6.14) with no body forces, (13.1) and (13.66) give us

$$\widehat{xx} + \widehat{yy} = \frac{2F}{\pi c} \left\{ \frac{1}{1+z^2/c^2} + \frac{1}{1+\bar{z}^2/c^2} - \sin^2 \alpha \right\},$$

and so $\widehat{xx} + \widehat{yy} \rightarrow -(2F/\pi c) \sin^2 \alpha$ as $|z| \rightarrow \infty$.

Similarly from (6.15), (13.5) and (13.66) give us

$$\widehat{xx} - \widehat{yy} + 2i\widehat{xy} = \frac{4F}{\pi c} \frac{(1 - z\bar{z}/c^2)}{(1 + \bar{z}^2/c^2)^2},$$

which $\rightarrow 0$ as $|z| \rightarrow \infty$. Comparing these results with (4.6), it is clear that we have a state of all-round tension T at infinity, where

$$T = -(2F/\pi c) \sin^2 \alpha.$$

This can be annulled, by adding the solution (11.22) in which T is taken as $(2F/\pi c) \sin^2 \alpha$, and the functions are taken for the variable z_1 , where

$$z_1 = z - d,$$

and relative to the origin $z_1 = 0$, which is the centre C of the hole, so that

$$c = d \tan \alpha, \quad d = a \cos \alpha, \quad c = a \sin \alpha, \quad (13.67)$$

where a is the radius of the hole. Using (13.67), this additive solution will be

$$\Omega_C(z_1) = (2Fc/\pi a^2) z_1, \quad \omega_C(z_1) = (4Fc/\pi) \log z_1.$$

For the origin O , we find the complex potentials are given, from (9.1), as

$$\Omega(z) = (2Fc/\pi a^2) z_1, \quad \omega(z) = (2F/\pi) \{2c \log z_1 - dc z_1/a^2\}. \quad (13.68)$$

Combining the complex potentials (13.66) and (13.68), the solution for the infinite plate under no body forces and containing a circular hole, unstressed save for the local application of the forces at $z = \pm ic$, is given by

$$\Omega(z) = \frac{2F}{\pi} \zeta, \quad \omega(z) = \frac{2F}{\pi} \left\{ \zeta z + 2c \log z_1 - \frac{dc}{a^2} z_1 \right\}, \quad (13.69)$$

where $z_1 = z - d$, after dropping an irrelevant constant from $\Omega(z)$. Bricas (1937, p. 231) gives the solution of this problem for the particular case in which $d = 0$, i.e. when the forces are at the ends of a diameter.

14. SOLUTIONS IN CYCLIC CO-ORDINATES

By cyclic co-ordinates we mean curvilinear co-ordinates ξ, η given by

$$\xi + i\eta = \zeta = 1/z. \quad (14.1)$$

[As a conformal transformation, this is an inversion followed by reflexion in the real axis; we could take the pure inversion $\zeta = 1/\bar{z}$, as an example of an isogonal (but not conformal) transformation defined by $\bar{z} = f(\zeta)$, and develop section 5 for such co-ordinates, but there appears to be little gain in so doing.]

Since $z\bar{z}(\zeta \pm \bar{\zeta}) = z \pm \bar{z}$,

the curves $\xi = \text{const.}$, $\eta = \text{const.}$, are families of circles given by

$$\left(x - \frac{1}{2\xi}\right)^2 + y^2 = \left(\frac{1}{2\xi}\right)^2, \quad x^2 + \left(y + \frac{1}{2\eta}\right)^2 = \left(\frac{1}{2\eta}\right)^2, \quad (14.2)$$

each family of circles touching one of the co-ordinate axes at the origin. We shall consider boundaries given by $2\eta = 1/r$ or $\bar{\zeta} = \zeta - i/r$.

In the case of gravitational body forces given by (3.3) and (3.5), (6.14) and (6.15) become

$$\widehat{xx} + \widehat{yy} = \frac{1}{2} \left\{ \Omega'(z) + \bar{\Omega}'(\bar{z}) + \frac{8\rho}{\kappa+1} (cz + \bar{c}\bar{z}) \right\}, \quad (14.3)$$

$$\widehat{xx} - \widehat{yy} + 2i\widehat{xy} = -\frac{1}{2} \left\{ z\bar{\Omega}''(\bar{z}) + \bar{z}\Omega''(z) - 4\rho \left(\frac{\kappa-1}{\kappa+1} \right) \bar{c}z \right\}, \quad (14.4)$$

where $c = -\frac{1}{2}ge^{-i\alpha}$.

Particular solutions \widehat{pq}_1 ($p, q = x, y, z$) for the stresses follow by taking the complex potentials

$$\Omega(z) = -4\rho cz^2/(\kappa+1), \quad \omega(z) = 0, \quad (14.5)$$

leading to $\widehat{xx}_1 + \widehat{yy}_1 = 0$, $\widehat{xx}_1 - \widehat{yy}_1 + 2i\widehat{xy}_1 = 2\rho\bar{c}z$,

or, in cyclic co-ordinates, using (6.17), (6.18) and (14.1)

$$\xi\bar{\xi}_1 + \eta\bar{\eta}_1 = 0, \quad \xi\bar{\xi}_1 - \eta\bar{\eta}_1 + 2i\xi\bar{\eta}_1 = \frac{\zeta^2}{\bar{\zeta}^2} (\widehat{xx}_1 - \widehat{yy}_1 + 2i\widehat{xy}_1) = 2\rho\bar{c} \frac{\zeta}{\bar{\zeta}^2}, \quad (14.6)$$

giving boundary stresses

$$4(\eta\bar{\eta}_1 + i\xi\bar{\xi}_1) = -4\rho\bar{c}\zeta/\bar{\zeta}^2. \quad (14.7)$$

From (3.9), (6.17), (6.18) and (14.1) complementary solutions $\widehat{p}q_0$ satisfy the equations for no body forces, i.e.

$$\xi\bar{\xi} + \eta\bar{\eta} = \widehat{xx} + \widehat{yy} = \frac{1}{2}\{\Omega'(z) + \bar{\Omega}'(\bar{z})\}, \quad (14.8)$$

$$\xi\bar{\xi} - \eta\bar{\eta} + 2i\xi\bar{\eta} = \frac{\xi^2}{\bar{\xi}^2}\{\widehat{xx} - \widehat{yy} + 2i\widehat{xy}\} = -\frac{\xi^2}{\bar{\xi}^2}\{z\bar{\Omega}''(\bar{z}) + \bar{\omega}''(\bar{z})\}, \quad (14.9)$$

the boundary stresses for the circular boundaries we consider here being given by

$$4(\eta\bar{\eta} + i\xi\bar{\eta}) = \Omega'(z) + \bar{\Omega}'(\bar{z}) + \frac{\xi^2}{\bar{\xi}^2}\{z\bar{\Omega}''(z) + \omega''(z)\}. \quad (14.10)$$

Consider the complex potentials

$$\Omega(z) = 4iA/z, \quad \omega(z) = 4iB \log z,$$

where A and B are real constants. Then from (14.10) we have

$$\eta\bar{\eta} + i\xi\bar{\eta} = i(A - B)\bar{\xi}^2 + 2iA\xi\bar{\xi} - iA\xi^2,$$

and if we make $B = 2A$ this becomes

$$\eta\bar{\eta} + i\xi\bar{\eta} = -iA(\xi - \bar{\xi})^2,$$

and along the circle $\bar{\xi} = \xi - i/r$, this makes

$$\eta\bar{\eta} = 0, \quad \xi\bar{\eta} = A/r^2. \quad (14.11)$$

From (7.4) the solution is non-dislocational in character, and from (8.7) and (8.9) the stress resultants for circles $\eta = \text{const.}$ are

$$X = 0, \quad Y = 0, \quad M = -2\pi A,$$

and the solution $\Omega(z) = -2i\frac{M}{\pi}\frac{1}{z}, \quad \omega(z) = -4i\frac{M}{\pi}\log z, \quad (14.12)$

is accordingly that for the space between two circles under no body forces, but uniform opposing shears over the two boundaries, which touch at $z = 0$, where the stresses become infinite. This is the limiting form of the solution (13.37), (13.38) as the real limiting points of the bipolar co-ordinates coalesce. We shall make use of it as a complementary solution in the following problem involving the gravitational body force.

Consider a heavy circular disk $r = a$ kept in equilibrium by a force and couple applied at the fixed point $z = 0$, the boundary being otherwise unstressed (see figure 9).

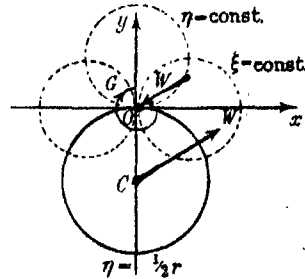


FIGURE 9

The fixing force is equal to the weight $W = \pi\rho a^2g$ in a direction $\alpha + \pi$, opposite to the direction of gravity, and the fixing couple is of amount $G = \pi\rho ga^3 \cos \alpha$ in the negative, i.e. clockwise, sense. Hence the force and couple resultants on the limiting *semicircle* of material about $z = 0$ are

$$X + iY = \pi\rho a^2ge^{i\alpha} = -2\pi\rho a^2\bar{c}, \quad M = \pi\rho a^3g \cos \alpha = -\pi\rho a^3(c + \bar{c}). \quad (14.13)$$

Now consider the complementary solution derived from the complex potentials, (B being a real constant),

$$\Omega(z) = A \log z + Bz, \quad \omega(z) = Cz \log z + Dz^2. \quad (14.14)$$

From (8.7) and (8.9) the stress resultants for the limiting *semicircle* in the material about $z = 0$ arising from (14.14) are

$$M = 0, \quad X + iY = \frac{1}{2}\pi(A - \bar{C}), \quad \text{hence take } C - \bar{A} = 8\rho a^2c. \quad (14.15)$$

If now we combine the particular solution given by (14.5) with the two complementary solutions given by (14.12) and (14.14), then from (14.7), (14.11) and (14.10) with (14.14), the complete boundary stresses are

$$4(\widehat{\eta\eta} + i\widehat{\xi\eta}) = 2B + A(\zeta - \bar{\zeta}) + 2i\frac{M}{\pi}(\zeta - \bar{\zeta})^2 + \bar{\zeta}\left(\bar{A} - 4\rho\frac{c}{\zeta^2}\right) + \bar{\zeta}^2\left(\frac{C}{\zeta} + \frac{2D}{\zeta^2}\right),$$

which becomes, along the boundary $\bar{\zeta} = \zeta - i/a$,

$$\begin{aligned} 4(\widehat{\eta\eta} + i\widehat{\xi\eta}) &= 2B + 2D - \frac{i}{a}(\bar{A} - A + 2C - 2\rho a^2c - 2\rho a^2\bar{c}) \\ &\quad + (C + \bar{A})\zeta - \frac{1}{a^2\zeta}(4\rho a^2c + C + 4iaD) - \frac{2}{a^2\zeta^2}(D - 2\rho iac), \end{aligned}$$

and we can make the boundary stresses vanish by choosing the constants so that the coefficients of the various powers of ζ vanish separately. Using (14.13) and (14.15), this necessitates

$$C = -\bar{A} = 4\rho a^2c, \quad B = \rho ga \sin \alpha, \quad D = 2\rho iac,$$

and the appropriate complex potentials are, since $c = -\frac{1}{2}ge^{-i\alpha}$,

$$\Omega(z) = \rho g \left\{ 2a^2e^{i\alpha} \log z + \frac{2}{\kappa + 1}e^{-i\alpha}z^2 + a \sin \alpha z - 2ia^3 \cos \alpha \frac{1}{z} \right\}, \quad (14.16)$$

$$\omega(z) = -\rho g \{ 4ia^3 \cos \alpha \log z + 2a^2e^{-i\alpha}z \log z + iae^{-i\alpha}z^2 \}. \quad (14.17)$$

When $\alpha = -\frac{1}{2}\pi$, we have the problem of a heavy disk resting on a horizontal plane at $z = 0$, the solution of which was first given by Michell (1900, p. 35). The generalized solution obtained here includes that in which the heavy disk is suspended by a string, given by taking $\alpha = \frac{1}{2}\pi$.

All the solutions we have found involving nuclei of strain where the displacements and sometimes the stresses become infinite at the nucleus, are physically unsuitable at the nucleus since the solutions are clearly outside the range of validity of a theory originally proceeding from the supposition that the displacements are small. But, in practice, ideal point loading cannot be achieved, and we imagine the solution for the practical loading spread over a small finite area to be closely imitated by the idealized mathematical problem, invoking Saint-Venant's principle to justify the neglect of the particular manner in which the stresses are applied to the boundary of the small region so long as the stress resultants for the small boundary are equivalent to the total loading on the small area.

SUMMARY

This paper gives an approach to two-dimensional isotropic elastic theory (plane strain and generalized plane stress) by means of the complex variable, resulting in a very marked economy of effort in the investigation of such problems as contrasted with the usual method by means of Airy's stress function and allied displacement function.

This is effected (i) by considering especially the transformation of two-dimensional stress; it emerges that the combinations $\widehat{xx} + \widehat{yy}$, $\widehat{xx} - \widehat{yy} + 2i\widehat{xy}$ are all-important in the treatment in terms of complex variables, (ii) by the introduction of two complex potentials $\Omega(z)$, $\omega(z)$ each a function of a single complex variable in terms of which the displacements and stresses can be very simply expressed.

Transformation of the cartesian combinations $u + iv$, $\widehat{xx} + \widehat{yy}$, $\widehat{xx} - \widehat{yy} + 2i\widehat{xy}$ to the orthogonal curvilinear combinations $u_\xi + iu_\eta$, $\widehat{\xi\xi} + \widehat{\eta\eta}$, $\widehat{\xi\xi} - \widehat{\eta\eta} + 2i\widehat{\xi\eta}$ is simple and speedy.

The nature of the complex potentials is discussed, and the condition that the solution for the displacements shall be physically admissible, i.e. single-valued or at most of the possible dislocational types, is found to relate the cyclic functions of the complex potentials.

Formulae are found for the force and couple resultants at the origin $z = 0$ equivalent to the stresses round a closed circuit in the elastic material, and these also are found to relate the cyclic functions of the complex potentials.

The body force has been supposed derivable from a particular body force potential which includes as special cases (i) the usual gravitational body force, (ii) the reversed mass accelerations or so-called 'centrifugal' body forces of steady rotation.

The power of the complex variable method is exhibited by finding the appropriate complex potentials for a very wide variety of problems, and whilst the main object of the present paper has been to extend the well-known usefulness of the complex variable method in non-viscous hydrodynamical theory to two-dimensional elasticity, solutions have been given to a number of new problems and corrections made to certain other previous solutions.

REFERENCES

- Bricas 1932 *Bull. Soc. Math. Grèce*, **13**, 11.
— 1935 *Bull. Soc. Math. Grèce*, **15**, 52.
— 1937 *La theorie de l'élasticité bidimensionnelle*. Athènes.
Cerothers 1920 *Proc. Roy. Soc. A*, **97**, 110.
Coker and Filon 1931 *Photoelasticity*. Camb. Univ. Press.
Edwardes 1893 *Quart. J. Math.* **26**, 270.
Ghosh 1926 *Bull. Calcutta Math. Soc.* p. 185.
— 1937 *Bull. Calcutta Math. Soc.* p. 177.
Hertz 1883 *Z. Math. Phys.* **28**, p. 125.
Inglis 1913*a* *Trans. Instn Nav. Archit., Lond.*, p. 219.
— 1913*b* *Engineering*, **95**, 415.
Jeffery 1921 *Phil. Trans. A*, **221**, 265.
Love 1927 *Mathematical theory of elasticity*, 5th ed. Camb. Univ. Press.
MacGregor 1935 *Trans. Amer. Math. Soc.* **38**, 177.
Melan 1932 *Z. angew. Math. Mech.* **12**, 343.
Michell 1899 *Proc. Lond. Math. Soc.* **31**, 100.
— 1900 *Proc. Lond. Math. Soc.* **32**, 35.
— 1901 *Proc. Lond. Math. Soc.* **34**, 134.
Milne-Thomson 1938 *Theoretical hydrodynamics*. Macmillan.
Timoshenko 1934 *Theory of elasticity*. McGraw Hill Book Co.

Stress systems in aeolotropic plates, III

By A. E. GREEN, PH.D. AND G. I. TAYLOR, F.R.S.

(Received 5 November 1940)

The stress distribution round a circular hole in an infinite aeolotropic plate subjected to tension in one direction is found theoretically in the case when the material of the plate has two directions of symmetry at right angles to one another. Numerical work is carried out using the elastic constants found in experiments made with specimens cut from the highly aeolotropic woods spruce and oak. An attempt is made to apply the calculated stress concentrations in conjunction with measurements of ultimate strength to determine the kind of failure that might be expected near a hole in a highly stressed wood plate.

INTRODUCTION

1. In previous papers (Green and Taylor 1939; Green 1939) fundamental stress functions were obtained for problems of generalized plane stress in a plate of aeolotropic material which has two directions of symmetry at right angles, and some of these functions were used to find the stresses produced by isolated forces acting in the plane of the plate. Numerical work was carried out for the highly aeolotropic materials oak and spruce, and the results were found to be in striking contrast with those for isotropic materials.

In the present paper the fundamental stress functions are used to find the stress distribution in an infinite aeolotropic tension member which contains a circular hole. These stress functions which satisfy the equations of equilibrium and which produce single-valued expressions for the corresponding stresses and displacements, are combined in an infinite series so as to satisfy the boundary conditions. The resulting formal solution is not completely satisfactory owing to difficulties of convergence in some parts of the plate. It may, however, be modified and expressed in a finite form which represents the stress distribution in the whole of the plate. As in previous papers, numerical work is carried out for certain specimens of oak and spruce.

FUNDAMENTAL EQUATIONS

2. Consider an elastic aeolotropic plate whose material has two directions of symmetry at right angles in the plane of the plate. The axes of x and y

are taken to be parallel to these directions and the z -axis will then be normal to the plate. The plate is imagined to be in a state of generalized plane stress in which the stress component $\hat{z}\hat{z}$ vanishes everywhere and the components $\hat{x}\hat{z}$ and $\hat{y}\hat{z}$ are zero at the surfaces of the plate. Then the *mean* stresses are given by

$$\hat{x}\hat{x} = \frac{\partial^2 \chi}{\partial y^2}, \quad \hat{y}\hat{y} = \frac{\partial^2 \chi}{\partial x^2}, \quad \hat{x}\hat{y} = -\frac{\partial^2 \chi}{\partial x \partial y}, \quad (2.1)$$

or, in polar co-ordinates,

$$\hat{r}\hat{r} = \frac{1}{r^2} \frac{\partial^2 \chi}{\partial \theta^2} + \frac{1}{r} \frac{\partial \chi}{\partial r}, \quad \hat{r}\hat{\theta} = -\frac{\partial}{\partial r} \left(\frac{1}{r} \frac{\partial \chi}{\partial \theta} \right), \quad \hat{\theta}\hat{\theta} = \frac{\partial^2 \chi}{\partial r^2}, \quad (2.2)$$

and the *mean* values of the stresses and strains are related by the equations

$$e_{xx} = s_{11}\hat{x}\hat{x} + s_{12}\hat{y}\hat{y}, \quad e_{yy} = s_{12}\hat{x}\hat{x} + s_{22}\hat{y}\hat{y}, \quad e_{xy} = s_{66}\hat{x}\hat{y}. \quad (2.3)$$

As shown in a previous paper (Green and Taylor 1939) the stress function χ satisfies the boundary conditions and the equation

$$\left(\frac{\partial^2}{\partial x^2} + \alpha_1 \frac{\partial^2}{\partial y^2} \right) \left(\frac{\partial^2}{\partial x^2} + \alpha_2 \frac{\partial^2}{\partial y^2} \right) \chi = 0, \quad (2.4)$$

where $\alpha_1 \alpha_2 = s_{11}/s_{22}, \quad \alpha_1 + \alpha_2 = (s_{66} + 2s_{12})/s_{22}, \quad (2.5)$

and the analysis is confined to cases where α_1 and α_2 are real and positive.

Three sets of plane polar co-ordinates are now introduced by the equations

$$x + iy = re^{i\theta}, \quad x + iy/\alpha_1^{\frac{1}{2}} = r_1 e^{i\theta_1}, \quad x + iy/\alpha_2^{\frac{1}{2}} = r_2 e^{i\theta_2}. \quad (2.6)$$

Then $r_1^{-2n} \cos 2n\theta_1, \quad r_2^{-2n} \cos 2n\theta_2, \quad \log r_1, \quad \log r_2, \quad (2.7)$

where n is a positive integer, are solutions of equation (2.4) which are symmetrical about both co-ordinate axes, and they give single-valued expressions for the corresponding stresses and displacements (Green and Taylor 1939). Expansions of these functions in terms of r and θ will be needed and are (Green and Taylor 1939)

$$\left. \begin{aligned} \frac{\cos 2n\theta_1}{r_1^{2n}} &= \frac{(1+\gamma_1)^{2n}}{r^{2n}} \sum_{s=n}^{\infty} \binom{n+s-1}{s-n} (-)^{s-n} \gamma_1^{s-n} \cos 2s\theta, \\ \log r_1 &= \log r + \sum_{s=1}^{\infty} \frac{(-)^{s-1} \gamma_1^s}{s} \cos 2s\theta, \end{aligned} \right\} \quad (2.8)$$

where

$$\gamma_1 = \frac{\alpha_1^{\frac{1}{2}} - 1}{\alpha_1^{\frac{1}{2}} + 1}, \quad (2.9)$$

with similar expansions for $r_2^{-2n} \cos 2n\theta_2$ and $\log r_2$ which are got by changing the suffix 1 into 2. By using the formulae (2.2) the corresponding stresses can be obtained as series which are absolutely and uniformly convergent.

THE TENSION PROBLEM

3. The plate is now supposed to contain a circular hole of radius a and the origin of co-ordinates is taken at the centre of the hole. A uniform tension T is then applied to the plate at infinity parallel to the x -axis. If the hole were absent the stresses would be derived from a stress function

$$\chi_0 = \frac{1}{2}Ty^2 = \frac{1}{4}Tr^2(1 - \cos 2\theta). \quad (3.1)$$

In order to allow for the effect of the hole suitable stress functions which give zero stresses at infinity are added to χ_0 so that the complete stress function gives zero normal and shear tractions over the boundary of the hole. The stress system must be single-valued, symmetrical about both the co-ordinate axes, and must give single-valued expressions for the corresponding displacements. Thus, it is assumed that the complete stress system may be derived from the stress function

$$\chi = \chi_0 + A_0 \log r_1 + B_0 \log r_2 + \sum_{n=1}^{\infty} \left\{ \frac{A_{2n} \cos 2n\theta_1}{(1 + \gamma_1)^{2n} r_1^{2n}} + \frac{B_{2n} \cos 2n\theta_2}{(1 + \gamma_2)^{2n} r_2^{2n}} \right\}, \quad (3.2)$$

where $A_0, A_1, \dots, B_0, B_1, \dots$ are constants which are to be found from the boundary conditions at the edge of the hole. Using (2.8) the stress function (3.2) may be expressed in terms of r and θ and then the corresponding stresses may be derived with the help of (2.2). The radial and tangential stresses are

$$\begin{aligned} \hat{r}\hat{r} = & \frac{1}{2}T(1 + \cos 2\theta) + \frac{A_0 + B_0}{r^2} + \frac{4}{r^2} \sum_{s=1}^{\infty} (-)^s s (A_0 \gamma_1^s + B_0 \gamma_2^s) \cos 2s\theta \\ & - \sum_{n=1}^{\infty} \sum_{s=n}^{\infty} \binom{n+s-1}{s-n} (4s^2 + 2n) (-)^{s-n} (A_{2n} \gamma_1^{s-n} + B_{2n} \gamma_2^{s-n}) \frac{\cos 2s\theta}{r^{2n+2}}, \end{aligned} \quad (3.3)$$

$$\begin{aligned} \hat{r}\hat{\theta} = & -\frac{1}{2}T \sin 2\theta + \frac{2}{r^2} \sum_{s=1}^{\infty} (-)^s (A_0 \gamma_1^s + B_0 \gamma_2^s) \sin 2s\theta \\ & - \sum_{n=1}^{\infty} \sum_{s=n}^{\infty} \binom{n+s-1}{s-n} (4sn + 2s) (-)^{s-n} (A_{2n} \gamma_1^{s-n} + B_{2n} \gamma_2^{s-n}) \frac{\sin 2s\theta}{r^{2n+2}}. \end{aligned} \quad (3.4)$$

If it is assumed that the expression (3.3) for $\hat{r}\hat{r}$ may be written as a cosine

series and that the expression (3.4) for $\bar{r}\bar{\theta}$ may be written as a sine series then the conditions for $\bar{r}\bar{r}$ and $\bar{r}\bar{\theta}$ to be zero when $r = a$ are

$$\left. \begin{aligned} \frac{1}{2}T + \frac{A_0 + B_0}{a^2} &= 0, \\ \frac{1}{2}T - \frac{4}{a^2}(A_0\gamma_1 + B_0\gamma_2) - \frac{6}{a^4}(A_2 + B_2) &= 0, \\ -\frac{1}{2}T - \frac{2}{a^2}(A_0\gamma_1 + B_0\gamma_2) - \frac{6}{a^4}(A_2 + B_2) &= 0, \end{aligned} \right\} \quad (3.5)$$

and

$$\left. \begin{aligned} \frac{4}{a^2}(-)^s(A_0\gamma_1^s + B_0\gamma_2^s) - \sum_{n=1}^s \frac{(-)^{s-n}}{a^{2n+2}}(4s^2 + 2n) \binom{n+s-1}{s-n} (A_{2n}\gamma_1^{s-n} + B_{2n}\gamma_2^{s-n}) &= 0, \\ \frac{2}{a^2}(-)^s(A_0\gamma_1^s + B_0\gamma_2^s) - \sum_{n=1}^s \frac{(-)^{s-n}}{a^{2n+2}}2s(2n+1) \binom{n+s-1}{s-n} (A_{2n}\gamma_1^{s-n} + B_{2n}\gamma_2^{s-n}) &= 0, \end{aligned} \right\} \quad (3.6)$$

for $s \geq 2$. These equations are satisfied by

$$\left. \begin{aligned} A_{2n} &= -\frac{Ta^{2n+2}\gamma_1^n(1+\gamma_2)(2n-1)!}{2(\gamma_1-\gamma_2)n!(n+1)!}, & B_{2n} &= \frac{Ta^{2n+2}\gamma_2^n(1+\gamma_1)(2n-1)!}{2(\gamma_1-\gamma_2)n!(n+1)!}, & (n \geq 1) \\ A_0 &= \frac{Ta^2(1+\gamma_2)}{2(\gamma_1-\gamma_2)}, & B_0 &= -\frac{Ta^2(1+\gamma_1)}{2(\gamma_1-\gamma_2)}. \end{aligned} \right\} \quad (3.7)$$

FURTHER DISCUSSION OF THE SOLUTION

4. A formal solution of the problem is given in the previous paragraph but it cannot be regarded as completely satisfactory because it is found that when $r = a$ the series (3.2) and the corresponding series (3.3) and (3.4) for the stresses do not converge for all values of θ , and for all values of γ_1 and γ_2 which lie between 0 and 1. Before writing down the boundary conditions the order of summations in (3.3) and (3.4) was reversed, and closer investigation shows that the resulting cosine series for $\bar{r}\bar{r}$ and sine series for $\bar{r}\bar{\theta}$ do actually converge for all values of θ , for all $r \geq a$ and for all values of γ_1 and γ_2 between 0 and 1. The series (3.2), when it converges, may be summed in finite terms and it will be seen that this sum gives a stress function which represents the complete solution of the problem for all the values of θ and r that are needed. Moreover, this sum may be expanded as a convergent cosine series from which may be derived a convergent cosine series for $\bar{r}\bar{r}$ and a convergent sine series for $\bar{r}\bar{\theta}$, these being the same as those used above but which were there obtained by a method which was only valid for some values of θ and r .

Instead of summing the series (3.2) it is easier to sum the series for the stresses which are derived from (3.2). It will then be found that the stresses are the *real* parts of the following expressions:

$$\begin{aligned} \bar{r} = & \frac{(1+\gamma_1)(1+\gamma_2)}{4\gamma_1\gamma_2} T \cos 2\theta \\ & + \frac{T}{4(\gamma_1-\gamma_2)} \left\{ \frac{(1+\gamma_2)(1-\gamma_1 e^{2i\theta})^2}{\gamma_1 e^{2i\theta} \left[1 - \frac{4a^2\gamma_1 e^{2i\theta}}{r^2(1+\gamma_1 e^{2i\theta})^2} \right]^{\frac{1}{2}}} - \frac{(1+\gamma_1)(1-\gamma_2 e^{2i\theta})^2}{\gamma_2 e^{2i\theta} \left[1 - \frac{4a^2\gamma_2 e^{2i\theta}}{r^2(1+\gamma_2 e^{2i\theta})^2} \right]^{\frac{1}{2}}} \right\}, \end{aligned} \quad (4.1)$$

$$\begin{aligned} \bar{\theta} = & -\frac{(1+\gamma_1)(1+\gamma_2)}{4\gamma_1\gamma_2} T \sin 2\theta \\ & - \frac{iT}{4(\gamma_1-\gamma_2)} \left\{ \frac{(1+\gamma_2)(1-\gamma_1^2 e^{4i\theta})}{\gamma_1 e^{2i\theta} \left[1 - \frac{4a^2\gamma_1 e^{2i\theta}}{r^2(1+\gamma_1 e^{2i\theta})^2} \right]^{\frac{1}{2}}} - \frac{(1+\gamma_1)(1-\gamma_2^2 e^{4i\theta})}{\gamma_2 e^{2i\theta} \left[1 - \frac{4a^2\gamma_2 e^{2i\theta}}{r^2(1+\gamma_2 e^{2i\theta})^2} \right]^{\frac{1}{2}}} \right\}, \end{aligned} \quad (4.2)$$

$$\begin{aligned} \bar{\theta}\theta = & -\frac{(1+\gamma_1)(1+\gamma_2)}{4\gamma_1\gamma_2} T \cos 2\theta \\ & - \frac{T}{4(\gamma_1-\gamma_2)} \left\{ \frac{(1+\gamma_2)(1+\gamma_1 e^{2i\theta})^2}{\gamma_1 e^{2i\theta} \left[1 - \frac{4a^2\gamma_1 e^{2i\theta}}{r^2(1+\gamma_1 e^{2i\theta})^2} \right]^{\frac{1}{2}}} - \frac{(1+\gamma_1)(1+\gamma_2 e^{2i\theta})^2}{\gamma_2 e^{2i\theta} \left[1 - \frac{4a^2\gamma_2 e^{2i\theta}}{r^2(1+\gamma_2 e^{2i\theta})^2} \right]^{\frac{1}{2}}} \right\}. \end{aligned} \quad (4.3)$$

Referred to cartesian axes the components of stress are the *real* parts of

$$\begin{aligned} \bar{x}x = & \frac{(1+\gamma_1)(1+\gamma_2)}{4\gamma_1\gamma_2} T \\ & + \frac{T}{4(\gamma_1-\gamma_2)} \left\{ \frac{(1+\gamma_2)(1-\gamma_1)^2}{\gamma_1 \left[1 - \frac{4a^2\gamma_1 e^{2i\theta}}{r^2(1+\gamma_1 e^{2i\theta})^2} \right]^{\frac{1}{2}}} - \frac{(1+\gamma_1)(1-\gamma_2)^2}{\gamma_2 \left[1 - \frac{4a^2\gamma_2 e^{2i\theta}}{r^2(1+\gamma_2 e^{2i\theta})^2} \right]^{\frac{1}{2}}} \right\}, \end{aligned} \quad (4.4)$$

$$\begin{aligned} \bar{y}y = & -\frac{(1+\gamma_1)(1+\gamma_2)}{4\gamma_1\gamma_2} T \\ & - \frac{T}{4(\gamma_1-\gamma_2)} \left\{ \frac{(1+\gamma_2)(1+\gamma_1)^2}{\gamma_1 \left[1 - \frac{4a^2\gamma_1 e^{2i\theta}}{r^2(1+\gamma_1 e^{2i\theta})^2} \right]^{\frac{1}{2}}} - \frac{(1+\gamma_1)(1+\gamma_2)^2}{\gamma_2 \left[1 - \frac{4a^2\gamma_2 e^{2i\theta}}{r^2(1+\gamma_2 e^{2i\theta})^2} \right]^{\frac{1}{2}}} \right\}, \end{aligned} \quad (4.5)$$

$$\bar{x}y = -\frac{iT}{4(\gamma_1-\gamma_2)} \left\{ \frac{(1+\gamma_2)(1-\gamma_1^2)}{\gamma_1 \left[1 - \frac{4a^2\gamma_1 e^{2i\theta}}{r^2(1+\gamma_1 e^{2i\theta})^2} \right]^{\frac{1}{2}}} - \frac{(1+\gamma_1)(1-\gamma_2^2)}{\gamma_2 \left[1 - \frac{4a^2\gamma_2 e^{2i\theta}}{r^2(1+\gamma_2 e^{2i\theta})^2} \right]^{\frac{1}{2}}} \right\}. \quad (4.6)$$

The square roots in the above formulae are to be evaluated by the following rule—if

$$R_1 e^{i\phi_1} = 1 - \frac{4a^2 \gamma_1 e^{2i\theta}}{r^2(1 + \gamma_1 e^{2i\theta})^2}, \quad R_2 e^{i\phi_2} = 1 - \frac{4a^2 \gamma_2 e^{2i\theta}}{r^2(1 + \gamma_2 e^{2i\theta})^2},$$

then ϕ_1 and ϕ_2 are to be chosen so that they lie between $-\pi$ and π , and R_1 and R_2 are positive. The square roots are then $R_1^{\frac{1}{2}} e^{i\phi_1}$, $R_2^{\frac{1}{2}} e^{i\phi_2}$.

It may easily be verified that this stress system satisfies the equations of equilibrium, gives zero values for $\hat{r}\hat{r}$ and $\hat{r}\hat{\theta}$ at the edge of the hole and reduces to a uniform tension T at infinity parallel to the x -axis. It therefore represents the complete solution of the problem, the form (3.2) being a valid expansion of the stress function for only a restricted range of values of θ and r . The stresses in an isotropic tension member containing a circular hole may be deduced by finding the limit of the above results as $\gamma_1 \rightarrow \gamma_2 \rightarrow 0$.

The stresses at the edge of the circular hole are of special interest and they take a comparatively simple form. Thus, when $r = a$,

$$\hat{\theta}\hat{\theta} = \frac{T(1 + \gamma_1)(1 + \gamma_2)(1 + \gamma_1 + \gamma_2 - \gamma_1\gamma_2 - 2\cos 2\theta)}{(1 + \gamma_1^2 - 2\gamma_1 \cos 2\theta)(1 + \gamma_2^2 - 2\gamma_2 \cos 2\theta)}, \quad (4.7)$$

$$\hat{x}\hat{y} = -\frac{1}{2}\hat{\theta}\hat{\theta} \sin 2\theta, \quad \hat{x}\hat{x} = \hat{\theta}\hat{\theta} \sin^2 \theta, \quad \hat{y}\hat{y} = \hat{\theta}\hat{\theta} \cos^2 \theta. \quad (4.8)$$

NUMERICAL DISCUSSION

5. In the first paper of this series numerical work was carried out for the problem of an isolated force acting at a point in a wide board of oak and of spruce cut so that the annular layers are parallel to the plane of the board. The elastic constants which were used for this purpose are reproduced in table 1, and these values are used for numerical work in the present paper. The grain of the wood will then be parallel to the y -axis or perpendicular to the tension. If the tension is applied parallel to the grain it is only necessary to interchange the values of s_{11} and s_{22} which will change the signs of γ_1 and γ_2 . The inverses of the constants s_{11} , ..., s_{66} are the Young's moduli. They are measured in kg./sq. mm.

TABLE 1

	s_{11}	s_{22}	s_{12}	s_{66}	α_1	α_2	γ_1	γ_2
Oak	10.15	1.72	-0.87	12.8	5.321	1.109	0.395	0.026
Spruce	15.5	0.587	-0.33	11.5	16.91	1.56	0.608	0.111

The values of the stresses on the edge of the circular hole have been evaluated by using the formulae (4.7) and (4.8). The results are given in tables 2 and 3. The stresses on the lines $\theta = 0$ and $\theta = \frac{1}{2}\pi$ which may be

obtained from (4.4) and (4.5) have also been evaluated and these results are reproduced in tables 4 and 5. In all the tables $\theta = 0$ is a line through the centre of the circular hole parallel to the applied tension.

TABLE 2. VALUES OF STRESSES ON THE EDGE OF THE CIRCLE: OAK

θ°	\widehat{xy}/T	\widehat{xx}/T	\widehat{yy}/T	$\widehat{\theta\theta}/T$
Tension parallel to grain				
0	0	0	-0.412	-0.412
10	0.0664	-0.0117	-0.376	-0.388
20	0.0996	-0.0363	-0.274	-0.310
30	0.0691	-0.0399	-0.120	-0.160
40	-0.0492	0.0413	0.0587	0.100
50	-0.263	0.313	0.221	0.534
60	-0.540	0.935	0.312	1.25
70	-0.752	2.07	0.274	2.34
75	-0.752	2.81	0.201	3.01
80	-0.626	3.55	0.110	3.66
90	0	4.36	0	4.36
Tension perpendicular to grain				
0	0	0	-2.43	-2.43
10	0.291	-0.0513	-1.65	-1.70
20	0.105	-0.0382	-0.289	-0.327
30	-0.343	0.198	0.595	0.793
40	-0.742	0.622	0.884	1.51
45	-0.873	0.873	0.873	1.75
50	-0.950	1.13	0.797	1.93
60	-0.940	1.63	0.543	2.17
70	-0.739	2.03	0.269	2.30
80	-0.404	2.29	0.0712	2.36
90	0	2.38	0	2.38

TABLE 3. VALUES OF STRESSES ON THE EDGE OF THE CIRCLE: SPRUCE

θ°	\widehat{xy}/T	\widehat{xx}/T	\widehat{yy}/T	$\widehat{\theta\theta}/T$
Tension parallel to grain				
0	0	0	-0.195	-0.195
10	0.0323	-0.0057	-0.183	-0.189
20	0.0543	-0.0198	-0.149	-0.169
30	0.0533	-0.0308	-0.0922	-0.123
40	0.0137	-0.0115	-0.0164	-0.0279
50	-0.0852	0.102	0.0715	0.173
60	-0.266	0.461	0.154	0.615
70	-0.529	1.45	0.193	1.65
75	-0.653	2.44	0.175	2.61
80	-0.683	3.87	0.120	3.89
85	-0.483	5.52	0.0423	5.57
90	0	6.37	0	6.37

TABLE 3 (continued)

θ°	\widehat{xy}/T	\widehat{xx}/T	\widehat{yy}/T	$\widehat{\theta\theta}/T$
Tension perpendicular to grain				
0	0	0	-5.14	-5.14
10	0.380	-0.0669	-2.15	-2.22
20	-0.187	0.0680	0.513	0.581
30	-0.735	0.424	1.27	1.70
40	-1.02	0.854	1.21	2.07
45	-1.06	1.06	1.06	2.13
50	-1.06	1.26	0.888	2.15
60	-0.922	1.60	0.533	2.13
70	-0.671	1.85	0.244	2.09
80	-0.352	2.00	0.0621	2.06
90	0	2.04	0	2.04

TABLE 4. VALUES OF STRESSES ON AXIS PERPENDICULAR TO TENSION

r/a	Spruce		Oak	
	\widehat{xx}/T	\widehat{yy}/T	\widehat{xx}/T	\widehat{yy}/T
Tension parallel to grain				
1.0	6.37	0	4.36	0
1.05	3.58	0.126	3.28	0.139
1.1	2.70	0.145	2.70	0.196
1.2	1.99	0.139	2.07	0.226
1.3	1.67	0.123	1.75	0.220
1.4	1.50	0.107	1.56	0.204
1.5	1.39	0.0937	1.44	0.186
1.6	1.32	0.0824	1.35	0.168
1.7	1.26	0.0730	1.29	0.152
1.8	1.22	0.0650	1.25	0.138
1.9	1.19	0.0583	1.21	0.125
2.0	1.17	0.0526	1.18	0.114
2.5	1.09	0.0337	1.10	0.0751
Tension perpendicular to grain				
1.0	2.04	0	2.38	0
1.05	1.97	0.096	2.24	0.116
1.1	1.90	0.182	2.11	0.205
1.2	1.77	0.325	1.91	0.340
1.3	1.67	0.436	1.76	0.431
1.4	1.59	0.522	1.64	0.491
1.5	1.52	0.589	1.54	0.527
1.6	1.45	0.640	1.46	0.548
1.7	1.40	0.678	1.39	0.557
1.8	1.36	0.705	1.34	0.557
1.9	1.32	0.724	1.30	0.552
2.0	1.29	0.737	1.26	0.543
2.1	1.26	0.743	—	—
2.2	1.23	0.745	—	—
2.3	1.21	0.744	—	—
2.4	1.19	0.740	—	—
2.5	1.18	0.733	1.14	0.486
3.0	1.11	0.677	—	—

TABLE 5. VALUES OF STRESSES ON AXIS PARALLEL TO TENSION

r/a	Spruce		Oak	
	\widehat{xx}/T	\widehat{yy}/T	\widehat{xx}/T	\widehat{yy}/T
Tension parallel to grain				
1.0	0	-0.195	0	-0.412
1.05	-0.0093	-0.176	-0.0160	-0.357
1.1	-0.015	-0.160	-0.0252	-0.310
1.2	-0.021	-0.132	-0.0270	-0.233
1.3	-0.019	-0.109	-0.0128	-0.176
1.4	-0.012	-0.0896	0.0116	-0.132
1.5	-0.0005	-0.0738	0.0425	-0.0985
1.6	0.014	-0.0606	0.0772	-0.0723
1.7	0.0309	-0.0496	0.114	-0.0521
1.8	0.0496	-0.0403	0.152	-0.0362
1.9	0.0698	-0.0325	0.190	-0.0237
2.0	0.091	-0.026	0.227	-0.0139
2.5	0.201	-0.0052	0.393	0.0113
3.0	0.307			
5.0	0.611			
10.0	0.871			
Tension perpendicular to grain				
1.0	0	-5.14	0	-2.43
1.05	-0.0311	-1.86	-0.0537	-1.40
1.1	0.0459	-0.926	-0.0281	-0.875
1.2	0.207	-0.270	0.0771	-0.370
1.3	0.337	-0.0498	0.188	-0.150
1.4	0.439	0.0399	0.287	-0.0407
1.5	0.518	0.0789	0.372	0.0165
1.6	0.582	0.0950	0.443	0.0477
1.7	0.634	0.100	0.504	0.0640
1.8	0.676	0.0998	0.555	0.0723
1.9	0.711	0.0967	0.599	0.0761
2.0	0.741	0.0920	0.637	0.0770
2.5	0.837	0.0679	0.764	0.0663
3.0	0.889			

The distribution of $\widehat{\theta\theta}$ over one quadrant of the hole is shown in figure 1 for the case where the tension is parallel to the grain and in figure 2 for the case when it is perpendicular to the grain. The distribution of $\widehat{\theta\theta}$ for a hole in an isotropic sheet, namely

$$\widehat{\theta\theta} = T(1 - 2 \cos 2\theta),$$

is also shown in figures 1 and 2 for comparison. In these figures the sheet is supposed to be in a state of tension in the direction $\theta = 0$ and $\widehat{\theta\theta}$ is positive where there is tension and negative where the stress is compressive. It will

be seen that with the highly aeolotropic spruce the maximum stress rises to $6.37T$ when the tension is applied parallel to the grain but that the tensile stress rises only to $2.04T$ when the tension is applied perpendicular to the grain. On the other hand, in the latter case there is a region where the compressive stress rises to $5.14T$. The regions of high stress concentration extend only over a small area however. It will be seen in figure 1 that the region where the stress exceeds that for an isotropic sheet extends only

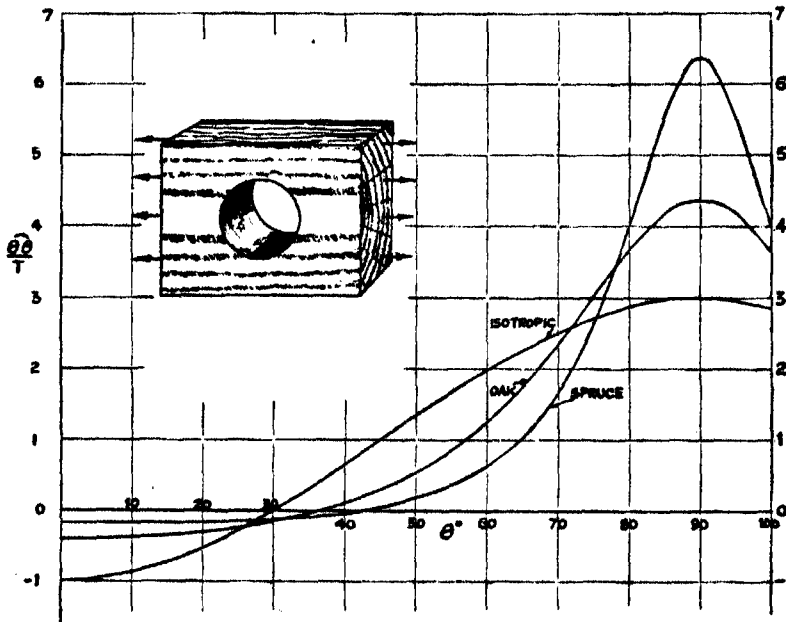


FIGURE 1. Stress distribution round the edge of a hole. Stress applied parallel to the grain, parallel to $\theta = 0$.

14 degrees from the position of maximum stress concentration. Similarly this region extends radially only 0.15 times the radius from the edge of the hole into the material. The distribution of $\theta\theta$ (or $\bar{x}\bar{x}$) along the y -axis is shown in figure 3.

It will be seen that with highly aeolotropic materials like wood the region of high stress concentration is limited to a small area where the fibres which have been cut in making the hole lie close to the uncut fibres.

The distribution of $\bar{x}\bar{y}$ round the edge of the hole is shown in figure 4.

The high stresses which occur near a hole in a stressed sheet have a technical interest for they may cause failure of a material which would otherwise have withstood the stress. In an isotropic sheet subject to tension

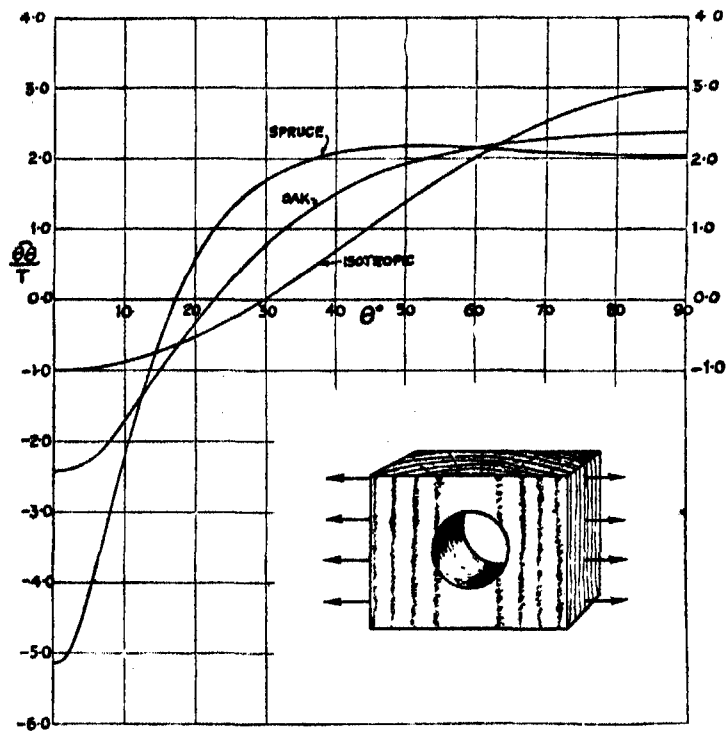


FIGURE 2. Stress distribution round the edge of a hole. Stress applied perpendicular to the grain, parallel to $\theta = 0$.

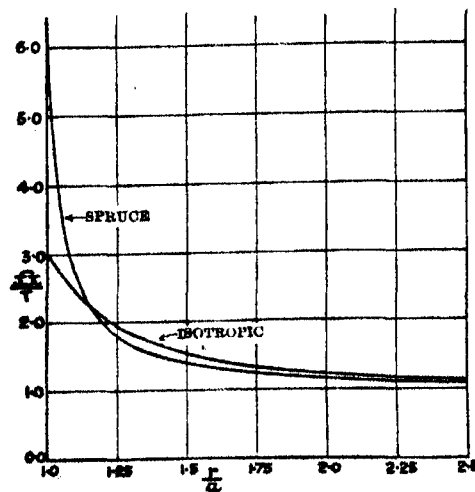


FIGURE 3. Distribution of $\hat{\sigma}_{\theta\theta}$ along $\theta = \frac{1}{2}\pi$ when stress is applied parallel to the grain and to $\theta = 0$.

in one direction a circular hole increases stress at the point of maximum stress concentration by a factor 3. Since the stress at the edge of the hole is a pure tension or compression in the circumferential direction, and the maximum stress occurs in the part of the circumference where the stress is tensile, failure of the material might be expected to be of the same type as that which occurs in the absence of the hole, but at a load only one-third as great. In actual practice the material usually withstands greater loads than this, a fact which has been explained by supposing that the stress at the point of maximum stress concentration does not reach its calculated value owing to slight plastic yielding there.

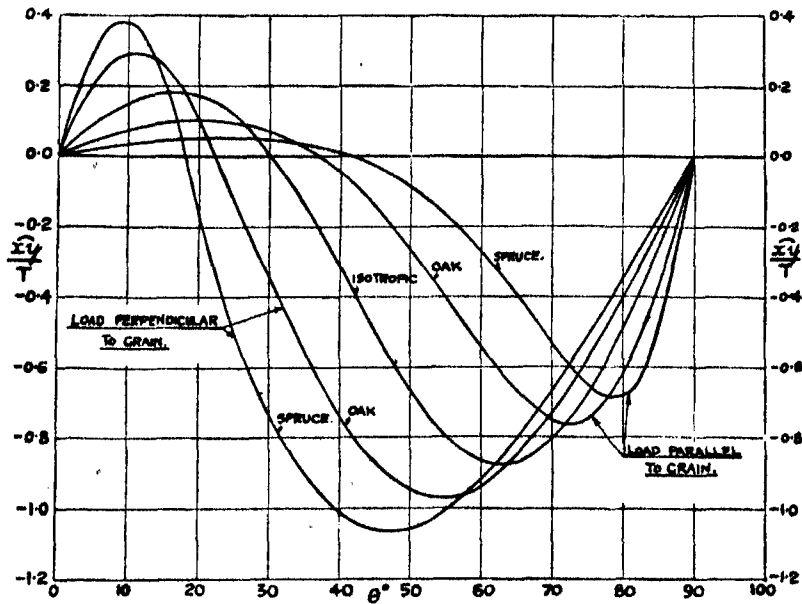


FIGURE 4. Distribution of shear stress $\hat{x}\hat{y}$ round the edge of a hole. Stress applied parallel to $\theta = 0$.

In anisotropic materials the conditions of failure under complex stresses have not been fully investigated, but the ultimate stresses for direct tensile and compressive loads applied parallel and perpendicular to the grain of certain woods as well as the ultimate shear stresses for shears applied in these directions have been measured. Measurements of this type are included in a report by C. F. Jenkin (1920) and the relevant figures* for spruce taken from this report are given in table 6.

* These figures, like those given in table 1 for the elastic constants, are average figures taken from results of experiments with a large number of specimens which exhibit considerable variations among themselves.

It is of interest to compare the loads at which failure of the five types contemplated in table 6 might be expected to take place in the neighbourhood of a hole in a spruce plank under tensile or compressive loads. For this purpose the value of the externally applied tensile or compressive stress, T , which would theoretically raise the stress near the hole to its ultimate value has been calculated. Four cases are considered:

- (i) tension applied parallel to the grain,
- (ii) compression applied parallel to the grain,
- (iii) tension applied perpendicular to the grain,
- (iv) compression applied perpendicular to the grain,

TABLE 6. ULTIMATE YIELD STRESSES FOR SPRUCE

	lb./sq. in.
Tension parallel to the longitudinal fibres	18,000
Compression parallel to the fibres (Corresponding in the present work with $\hat{x}\hat{x}$ when the external load is applied along the grain and $\hat{y}\hat{y}$ when it is perpendicular to it)	5,000
Tension acting in the tangential direction, i.e. perpendicular to a vertical plane through the centre of the tree when standing vertically	400 to 800
Compression in this direction (Corresponding with $\hat{y}\hat{y}$ when the load is applied along the grain and $\hat{x}\hat{x}$ when applied perpendicular to it)	700
Shear stress across a plane perpendicular to the tangential direction (Corresponding with $\hat{x}\hat{y}$)	1,100

and for each case the value of T is calculated for each of the five types of failure referred to in table 6 namely

- (a) rupture of the longitudinal fibres,
- (b) breakdown of longitudinal fibres in compression,
- (c) rupture perpendicular to the fibres,
- (d) breakdown by compression perpendicular to the fibres,
- (e) shearing rupture parallel to the fibres.

Taking case (i), (a) the stress maximum $\hat{x}\hat{x}$ occurs at $\theta = 90^\circ$ and is equal to $6.37T$. From table 6 this corresponds with 18,000 lb. so that the tensile load necessary to cause breakdown of type (a) in case (i) is

$$T = 18,000/6.37 = 2800 \text{ lb./sq. in.}$$

In each case the appropriate maximum positive or negative values of $\hat{x}\hat{x}$, $\hat{y}\hat{y}$, $\hat{x}\hat{y}$ have been taken from table 3. The results are given in table 7.

TABLE 7. CALCULATED STRESSES FOR FAILURE
NEAR A HOLE IN A SPRUCE PLANK

Failure type lb./sq. in.	Case (i)	Case (ii)
a	$\frac{18,000}{6.37} = 2800$	$\frac{18,000}{0.0308} = 580,000$
b	$\frac{5000}{0.0308} = 160,000$	$\frac{5000}{6.37} = \boxed{780}$
c	$\frac{400 \text{ to } 800}{0.193} = 2000 \text{ to } 4000$	$\frac{400 \text{ to } 800}{0.195} = 2000 \text{ to } 4000$
d	$\frac{700}{0.195} = 3600$	$\frac{700}{0.193} = 3600$
e	$\frac{1100}{0.683} = \boxed{1600}$	$\frac{1100}{0.683} = 1600$
	Case (iii)	Case (iv)
a	$\frac{18,000}{1.27} = 14,000$	$\frac{18,000}{5.14} = 3500$
b	$\frac{5000}{5.14} = 970$	$\frac{5000}{1.27} = 3900$
c	$\frac{400 \text{ to } 800}{2.04} = \boxed{200 \text{ to } 400}$	$\frac{400 \text{ to } 800}{0.067} = 6000 \text{ to } 12,000$
d	$\frac{700}{0.067} = 10,000$	$\frac{700}{2.04} = \boxed{350}$
e	$\frac{1100}{1.08} = 1000$	$\frac{1100}{1.08} = 1000$

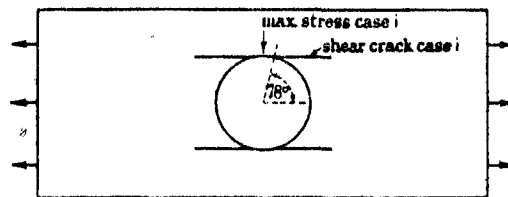


FIGURE 5a. Shear crack when stress is applied parallel to the grain

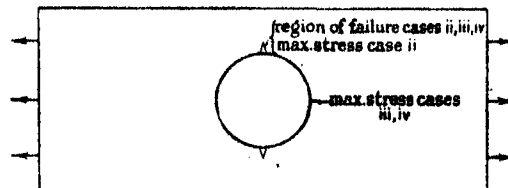


FIGURE 5b. Position of failure when compression is applied parallel to grain or when tension or compression is applied perpendicular to grain.

The type of failure which might be expected to occur corresponds with the least of the values given in table 7 for each of the four cases. These are indicated in table 7 by lines enclosing the four calculated values. It will be noticed that in cases (ii), (iii) and (iv) failure may be expected at the point on the circumference of the hole which lies on the diameter perpendicular to the direction of the applied stress. This is the position of maximum stress when the stressed sheet is isotropic. In case (ii) this position corresponds with the point of maximum stress concentration, but in cases (iii) and (iv) greater stresses occur in other parts of the field. Case (i) is more interesting. Here the type of breakdown by shearing parallel to the grain is quite different from that in an isotropic sheet. The maximum value of $\bar{x}y$ occurs at $\theta = 78^\circ$ (see figure 4). The calculated position of the shear crack is shown in figure 5*a*. The position of the region where breakdown is to be expected in cases (ii), (iii) and (iv) is shown in figure 5*b*.

REFERENCES

- Green, A. E. 1939 *Proc. Roy. Soc. A*, **173**, 173.
Green, A. E. and Taylor, G. I. 1939 *Proc. Roy. Soc. A*, **173**, 162.
Jenkin, C. F. 1920 *Report on Materials of Construction used in Aircraft and Aircraft Engines*. London: Stationery Office.

Short period fluctuations in the characteristics of wireless echoes from the ionosphere

By T. L. ECKERSLEY, B.A., B.Sc., F.R.S. AND
F. T. FARMER, B.Sc., Ph.D.

(Received 6 August 1940)

[Plates 6, 7]

A new experimental technique is described for studying the directional and polarization characteristics of wireless echoes. The method makes use of an ordinary spaced frame direction-finding system, combined with a twin-channel amplifier and cathode-ray tube, so that by observations of an ellipse on the screen the phases and amplitudes of the e.m.f.s in the aerial system may be determined. By appropriate connexions of the aerials either of the two characteristics mentioned may be studied.

The equipment is made applicable to pulse transmissions by means of an 'echo selecting' device, which has the function of desensitizing the apparatus for all but a small element of the time-base cycle. The duration of the active period is of the order of $15\mu\text{sec.}$, and by means of suitable phasing circuits it may be made to coincide with the arrival of any echo of the pulse pattern, causing the ellipses appropriate to that element to appear alone on the cathode-ray tube screen. In polarization measurements a single ellipse is produced on the screen corresponding to the e.m.f.s in two loops at right angles. In directional measurements, the aerials are switched at 25 cyc./sec. by relays so that two ellipses are produced, one representing the phase difference in one pair of spaced frames, and the other that in the other pair. These are separated laterally on the tube by a synchronized magnetic deflecting circuit.

The figures are photographed on a moving film, and the ellipses due to each pulse from the transmitter are thus recorded separately. In this way rapid changes in polarization and direction can be studied.

The method has been applied to the various types of ionospheric echo, and records are shown to illustrate their behaviour. In particular it has been used to study the short-lived scatter echoes obtained on high power transmissions, whose characteristics change too rapidly to be followed by manual methods. The results of both polarization and directional measurements are considered in relation to the general problem of ionospheric irregularities.

INTRODUCTION

A large amount of data concerning the properties of wireless signals reflected from the ionosphere is now available. This has been mainly derived by the pulse technique of Breit and Tuve, which, in conjunction with special receiving apparatus, has enabled measurements to be made of the various quantities relating to any individual echo, namely, amplitude,

time delay, polarization and direction of arrival. From such experiments our present knowledge of the ionosphere has been largely gained, and the main structure of the reflecting regions is now known with some certainty.

In many respects it is possible to interpret such experimental results in terms of uniformly stratified layers of ionization, which vary only slowly from one set of measurements to the next; and it is well known how fully the complicated behaviour of echoes with varying wave frequency may be explained in terms of this simple picture, when account is taken of the earth's magnetic field. The evidence that this is substantially a correct description of the ionosphere is, indeed, very great.

It has been realized for a long time, however, that this picture is inadequate to explain all the properties of down-coming waves, and that some degree of irregularity in the ionized layers is always present. The first evidence pointing to this was obtained from observations on signals within the skip zone of a transmitter, where, apart from a possible ground wave, no signal would be expected. These waves were shown to be due to sporadic reflexions from the ionosphere, and in a detailed study of this subject, Eckersley (1940) has shown that they have their origin mainly at the level of the E layer. Other experiments have been made on the variability of echoes from the E and F layers. The fading of these echoes on two spaced receivers has been studied by Ratcliffe and Pawsey (1933), who showed that the layers behave as if they were slightly distorted in structure, thereby giving interference beats between different components of a diffraction system; while Martyn and Green (1935) and others have used direction-finding apparatus to trace the variations in angle of arrival of pulse echoes, and have shown that for both E and F reflexions there are always appreciable departures from the mean direction that would be expected for mirror reflexion. Many other experiments, including polarization measurements, have been made, and have confirmed the view that irregularities play an important part in ionospheric propagation.

The data available hitherto have been obtained chiefly by manual methods; that is to say, directions of arrival have been measured by the ordinary technique in which a balance of any particular echo of a pulse pattern is made by turning a goniometer or a phase-adjusting condenser; and polarizations have been determined similarly by using an appropriate aerial system and a goniometer for balancing the echo considered. It has consequently been possible to study only the more slowly varying characteristics of the echoes.

For reflexions from the F region, the directions and polarizations generally change sufficiently gradually to meet this requirement; but in the case of

E reflexions, and more especially of the sporadic echoes mentioned above, this is by no means the case, and a detailed knowledge of their characteristics has not, therefore, been obtainable. There is reason to believe that the ionic 'clouds' responsible for these latter echoes also cause much of the irregularity observed on other reflexions, the deviations in direction of waves returned from the E or F regions, for instance, being due to their passage through a 'scattering zone' at this lower level; and a fuller knowledge of their rapidly changing characteristics is, therefore, of some importance. The present work is concerned with this problem, as well as with the more rapid changes that occur in E and F reflexions. •

In measurements by the manual method, the effect of scattering in the path of a wave is to render the balance obtainable on a pair of spaced frames imperfect. This implies that the phase surface is changing at such a rate that, in the time taken for adjustment of the apparatus, the balance position has changed appreciably, and only a mean value over the period involved can be derived. The conditions are strictly analogous to those of Michelson's optical interferometer, in which spaced mirrors take the place of the receiving aeriels; and the same relations govern the two types of measurement (Eckersley 1938). Thus by measuring the *degree* of balance on any signal, it is possible to determine a cone angle within which the component rays may be said to lie; this angle, of course, being a function of the time taken in the adjustment of the apparatus. In the earlier investigations referred to, this method was used extensively to determine the spread of the various echoes received; and it was found, in particular, that the type of scatter echo that is received within the skip zone shows a very big spread in angle, implying that the rays of which it is composed are very diverse in direction. It was not possible, however, to state how quickly the resultant phase surface changed, and so to what extent the time of measurement would have to be reduced to obtain a single value for the direction of this surface. The time taken in manual observations is of the order of two or three seconds; and all that can be stated is that during this time the echo behaves as if it were a bundle of waves of random phase relative to one another, and extending in angle over the cone calculated. The same considerations apply to measurements of polarization. A mean value over a period of time of the same order is all that can be specified. In the case of these echoes in particular, therefore, uncertainty persists as to the rapid fluctuations of polarization and direction that take place; and in the present work, the technique of more detailed measurement has been applied to these, as well as to other ionospheric reflexions, in order to resolve such uncertainties as far as possible.

GENERAL CONSIDERATIONS

Although, as has been stated, most of the work on echo directions and polarizations has been done by the manual technique, certain alternative methods have already been published, which must be considered briefly. We are here concerned only with those applicable to pulse transmissions, for which it is clear the problem of any automatic registration method is greatly increased.

Martyn, Piddington and Munro (1937), in Australia, have developed a method of recording polarizations of echoes, in which each pulse on the time base appears as an ellipse superimposed on the moving trace. The ellipses are produced by picking up the signal on two loops at right angles, amplifying the e.m.f.s separately, and, after rectifying, applying them to the pairs of plates of a cathode-ray tube. By superimposing a weak ground wave of slightly different frequency, the output voltages are made to oscillate in amplitude, and when the two amplifiers are put on together, an ellipse is traced out on the tube for each echo, representing the polarization of the wave concerned. Thus, by taking photographs in rapid succession, it is possible to record the changes of any one echo, even when they are taking place quickly. The method has been used very successfully by these authors, but it has drawbacks which appear to make it unsuitable for any detailed analysis of the type here contemplated. In the first place, it is incapable of resolving two echoes which are closely spaced on the time base, since the ellipses due to them must necessarily overlap; and moreover, the movement of the spot along the time base during the generation of an ellipse spreads the figure laterally, and so hinders accurate measurement of its shape. In addition, there is the difficulty that any one echo comprises e.m.f.s of all amplitudes up to the peak of the pulse, owing to the finite time taken for its growth and decay, and this tends to blur the pattern still further. These factors seem to limit the usefulness of the method considerably, and it was, therefore, not applied to the present problem.

A technique for observing instantaneously directions of echoes has been described by Barfield and Ross (1938). In this, spaced aeriels are used to pick up the signal, and the outputs are connected through twin-channel amplifiers to the plates of an oscillograph. The aeriels are arranged in such a way that the figure produced on the screen is always a straight line, and its orientation is a measure of the direction of arrival. The bearing can thus be read on a circular scale, and by photographing the figure repeatedly, changes in direction can be traced even when these occur rapidly. The method is essentially one for continuous wave direction-finding, and for

signals of this type well-defined figures are produced on the tube; but it has been applied to pulse transmissions also, for which a number of lines of different orientation are produced corresponding to the various echoes received. When the echo system is simple, and attention is directed to one of the stronger components, there is probably little difficulty in tracing the changes in any particular echo, with the help of a monitor receiver to identify the pulses on the tube, though even under these conditions there is likely to be some confusion due to overlapping of the figures. When a large number of echoes are received, however, or observations are required on a weak echo in the presence of much stronger ones, the limitations appear to be serious, and photographs of the complex figures on the tube would be very difficult to analyse. No attempt seems to have been made to apply the method to these more stringent conditions.

EXPERIMENTAL ARRANGEMENTS

The technique used in the present work was designed to overcome these difficulties as far as possible. In principle it consisted in isolating a small element of the group of echoes received, and confining observations of polarization and direction to that alone. In this way a thin section taken from any one echo could be studied without confusion by the rest of the signal, and changes in the characteristics relating to that element could be followed in much greater detail.

The same essential apparatus was used for both polarization and directional measurements. This comprised a spaced frame aerial system, such as has been used extensively for direction finding by the phase-balance method (Eckersley 1935), and a twin-channel amplifier connected to a cathode-ray tube. The figure on the latter determined the phase and amplitude relations between the e.m.f.s fed into the inputs of the amplifiers, and by suitable arrangements of the aeri^{al}s, these could be made to yield either of the two characteristics relating to the echo selected.

The aeri^{al}s consisted of four frames arranged as shown in figure 1, lying on the geographical axes through the centre of the observing hut. Each was aligned carefully relative to its feeder, so that spurious pick-up on this conductor should not be transferred to the frame; and thus by comparing the phases of the e.m.f.s in opposite members, the direction of arrival of any wave could be determined. In using such apparatus manually, a balance is made on a phase unit at the centre, first with one pair of frames connected, and then with the other; and from the two cone angles so obtained, the complete direction of arrival of the wave can be calculated.

In order to apply this same principle to the present experiments, an automatic switching system was introduced, so that the change-over could be made rapidly and repeatedly, and this was achieved by means of relays operating on the input connexions to the amplifiers, as shown in figure 2. It was arranged that the amplifiers should be coupled to one pair of frames for $1/50$ sec. and to the other pair for the next $1/50$, and so on, so that when transmissions controlled from the supply mains were used, alternate pulses were received on either of the two pairs of frames in turn. Simultaneously with this change-over, the cathode-ray tube beam was deflected sideways by magnetic coils, so that two separate ellipses were produced on the screen; one of these then indicated the phase relation between the e.m.f.s in the North-South pair, and the other that in the East-West pair of frames. In a time of $1/25$ sec., therefore, the complete direction of arrival was indicated on the tube.

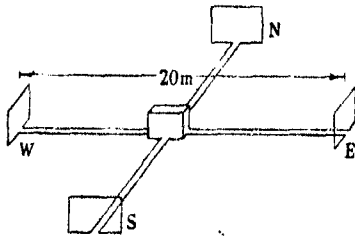


FIGURE 1. Spaced frame aerial system.

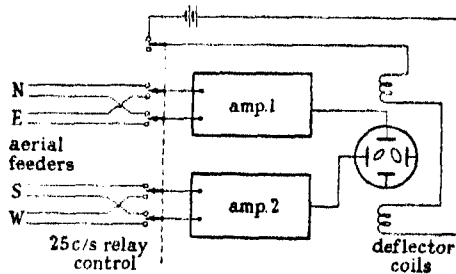


FIGURE 2. Aerial switching system.

For polarization measurements, the connexions were altered so that the frames of the North-South pair were coupled in parallel to one amplifier, and those of the East-West pair to the other, no automatic switching being in operation. It is easy to see that, for vertically incident waves, the two pairs in this way behave as a single pair of crossed loops at the centre of the system, such as would normally be used for polarization observations. They also had the advantage of being well displaced from the possible distortion of apparatus in the hut. For waves arriving obliquely, there was a phase difference between the frames in parallel, and this had to be allowed for in calculating the relative amplitudes of the two fields at right angles. It did not affect the phase relation of the resultant e.m.f.s. Provision was made for easy interchange between the two types of aerial connexion, so that either polarization or directional measurements could be made as desired.

A choice of two sets of frames was available, one having a spacing between opposite members of 20 m., and the other of 80 m. Normally, the closer spacing was used for polarization measurements, in which it was desirable to keep the extent of the aerial array small compared with a wave-length,* and the wider spacing was used for directional tests, where greater sensitivity was obtained by increasing the separation between frames.

The system used for selecting an element of any particular echo is shown diagrammatically in figure 3. The principle is similar to that described in a previous paper (Farmer 1935), in which the amplitudes of individual echoes were recorded automatically, but in the present case the selecting action was carried out by controlling the modulating electrode of the

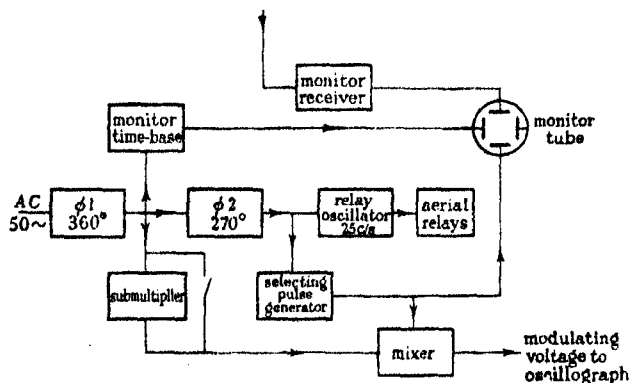


FIGURE 3. Echo selecting system (schematic diagram).

cathode-ray tube, instead of by varying the sensitivity of the receiver. Thus, for the whole time, the high-frequency output of the amplifiers was applied to the plates of the tube, but an image was formed on the screen only for the brief period in which the tube was caused to focus. The modulation of the tube was achieved by a circuit which generated a square-shaped positive pulse every 1/50 sec., this pulse being applied to the control anode through a condenser coupling. By adjusting the mean potential of this anode so that the tube was just 'blacked out' between pulses, it could be arranged that a correctly focused pattern should be produced during the short periods in which the impulse was operative, i.e. during a fraction of the time in which the echo to be observed was arriving. The extent of this period could be varied at will from about 10 to 200 μ sec.

* The wave-lengths used were of the order of 50 m.

Some elaboration arose from the necessity of making the phase of the selecting impulse variable with respect to the supply mains controlling the transmitter. This was required not only to allow a choice of the portion of the echo pattern selected, but to make the apparatus adaptable to transmitters having different phase relations to the A.C. supply. The arrangement of figure 3 met this purpose. From the incoming supply mains an impulse of variable phase was generated by the phase changer ϕ_1 and suitable output circuit, and this was used to lock a monitoring time base on which the echoes obtained on an independent receiver could be observed. Thus any transmitter could be made to give its echo pattern suitably displayed on this time base, and the start of the sweep provided a convenient datum relative to which the selecting impulse could then be adjusted. This was done by the second phase changer ϕ_2 which controlled the oscillograph modulating circuit just mentioned. Both phase changers had to work over a large range, and specially designed circuits were developed to provide smooth variation between the extreme positions.

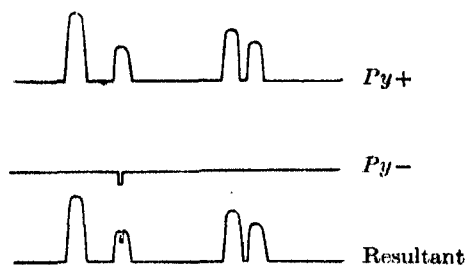


FIGURE 4. Monitor tube.

In order to show in which part of the echo pattern the selecting impulse was operative, a voltage similar to that used to modulate the main oscillograph was applied to one of the vertical deflecting plates of the monitor tube. A small deflexion was thus produced in the time base, as shown in figure 4, and by adjustment of the phase, this could be placed in the middle of any echo as desired. The width of the deflexion indicated the duration of the active period of the main tube.

A photographic method was used to record the ellipses generated by the echoes. This consisted of a simple camera in which a film was moved steadily past the field of view; by virtue of the oscillograph modulation, the image on the screen only existed for a few micro-seconds in every cycle, and no cinematograph mechanism was therefore necessary to separate the successive ellipses. It was found by using a tube with very small after-glow, that quite distinct images were obtained even when the film was

moved rapidly across the field. In order to obtain sufficient exposure, however, in the very short time (normally about 15 μ sec.) during which the tube was focused, a special photographic tube capable of working up to 10,000 V was necessary.*

With the repetition frequency of 50 per sec. of the successive pulses, it is obvious that a fast-moving film was necessary to separate the resulting ellipses. In the polarization tests, a speed of some 15 cm./sec. was required to obtain clear separation; in the directional tests each ellipse repeated at half the frequency of the pulses, and a somewhat slower speed was sufficient. In many cases, however, it was found that the direction or polarization did not change much in the short time between successive pulses, and the high film speed thus became wasteful. A device was, therefore, incorporated by which the number of ellipses appearing on the screen could be cut down to any desired extent. This took the form of an oscillator which operated at a subharmonic of the mains frequency, and which desensitized the oscillograph for all but a fraction of its normally active periods. It was designed so that the active part of the cycle (in directional experiments) should be exactly two periods of the mains supply, during which an ellipse on each of the two pairs of frames would be traced out, and that during the remainder of the time the tube should be idle, regardless of the length of the cycle. A multivibrator type of circuit was used for this purpose, and frequencies of 1/6, 1/8, 1/10 and 1/12 of the 50-cycle mains were employed, an even subharmonic being chosen to obtain regularly spaced ellipses on the film. In polarization experiments a slightly different condition was necessary, in that only one ellipse had to be admitted per cycle instead of two, and provision was made for interchange between the two types of operation by simple switching.

EXPERIMENTAL PROCEDURE

In either type of experiment careful lining-up of the apparatus was necessary before commencing.

In polarization measurements this was achieved by means of an aerial on a line at 45° to the axes of the frames, which was connected by a feeder to an oscillator in the hut. The two amplifiers were tuned carefully to the signal, and by means of line-up switches interconnecting their grids, the phases were equalized at successive stages in the two units. Finally, the input circuits were adjusted to give equal phase output from the calibrating

* A Cossor tube, type 3222, met this requirement very well.

aerial, as shown by a line at 45° on the tube. Since, as far as the input to the receivers was concerned, the signal from this aerial was equivalent to a plane polarized wave incident vertically on the system, it is obvious that when the circuits were thus lined up, the shape of the figure produced by a down-coming wave was a direct measure of the polarization of the wave on the ground; and if the wave was reflected vertically from the ionosphere, it gave at once the polarization ellipse of the wave front. In the case of waves incident obliquely, it was possible to calculate the polarization from the measured figure by allowing for the phase difference between opposite frames, as mentioned above, and then projecting the ellipse on the ground on to the wave front. This latter process was most readily done by the transformation method described by Eckersley and Millington (1939).

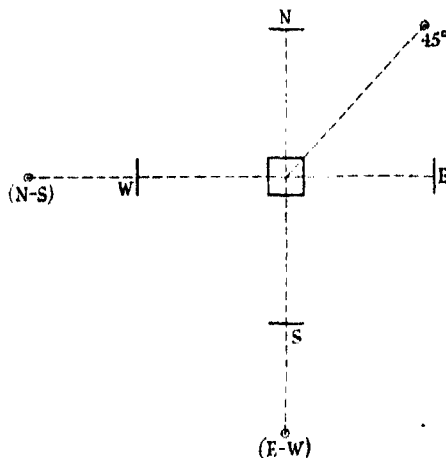


FIGURE 5. Lining-up aerials.

In directional experiments a different method of lining-up was required. Two aerials were used, one situated due West of the apparatus and the other due South, and these were used to line-up respectively the North-South and the East-West pairs of frames. The arrangement is shown in figure 5. The aerials were connected through switches to an oscillator in the hut, and the process of adjustment consisted in first tuning each amplifier in the way described for polarizations, and then making final adjustments to the input circuits, so that either aerial with its appropriate pair of frames gave a line at 45° on the oscillograph. Under these conditions the phase difference between the e.m.f.s in the frames, due to a down-coming wave, could at once be determined from the shape of the resulting ellipse; and

from the two figures together, the complete direction of arrival of the wave could be obtained.

Provision was made for determining the sense of rotation of the ellipses by connecting a small condenser across the tuned circuit in one of the i.f. amplifier stages. This was operated by hand, and it was, therefore, not possible to keep track of the sense of rotation when this latter was changing rapidly. In the case of echoes of this type, however, the variations in direction or polarization were the factors of chief interest, and the mean absolute values could be determined, when necessary, by the manual method.

DISCUSSION OF RESULTS

Pulse transmissions from three stations were used. These were Daventry, giving a weekly programme on 6.1 Mc./sec. during the sunrise period, Ongar (GOQ) giving a similar programme on 7.6 Mc./sec. from 0900 to 1100 G.M.T., and an experimental transmitter in Chelmsford which could be controlled from the receiving hut, and which gave a wide choice of frequencies. Of the three stations, the two former were of relatively high power, but Ongar alone, on 20 kW, gave sufficient radiation to show good scatter reflexions. The programmes from this station were, unfortunately, much curtailed by other requirements, and it has not been possible to get all the data that was desired on these echoes; the stronger reflexions, however, could be studied with lower power, and have been observed in greater detail.

We shall, for convenience, consider the two types of experiment separately.

(a) *Directional measurements*

(1) *Split F echoes.*

In the sunrise experiments on Daventry, the F region density was increasing with elevation of the sun, and it usually became sufficient to reflect the 6.1 Mc./sec. waves during the test. An opportunity was therefore provided for examining the separate magneto-ionic components from this region.

Figure 6 shows a typical record taken under these conditions, obtained by selecting an element from the first order F extraordinary echo. The duration of the selecting period was about 15 μ sec., and the submultiplier was set to a frequency 1/6 of that of the mains, so that at each cycle of this

frequency one pair of ellipses was recorded on the film. In this, as in some of the other experiments, the camera was turned by hand, resulting in slight irregularities in the spacing of the ellipses.

The most noticeable feature of the record is the constancy of shape of the ellipses throughout the period of 11 sec., although their amplitudes change as the signals fade. This is typical of single echoes from the F region, and in particular of a single magneto-ionic component; and it is interesting to relate it to the type of observations obtained manually by the phase-balance method. In experiments of that type, it was found that the variations in direction were nearly always slow enough to follow manually, and at any instant a good balance could be obtained. It would be expected, therefore, that when the speed of observing is increased, no appreciable variations should occur in the period of a few seconds in which a manual adjustment could be made. From measurements of the ellipses, it is found that the phase difference on either pair of frames did not change by more than 2° during the 11 sec. of record shown, and by a scarcely greater amount in the whole recording of 35 sec. Since the frame spacing was 80 m., and the waves were arriving nearly vertically, it is easy to see that the direction remained constant to within 0.2° during this period. At other times larger variations have occurred, corresponding, presumably, to the conditions when manual balances were less easily obtained, but in every case in which the rapid technique was used, and observations were made on a single magneto-ionic component, it has been found that the direction remained sensibly constant for at least a few seconds.

The actual direction of the waves may be calculated from the record, though unfortunately in these tests there was some doubt as to the sense of rotation of the ellipses. The calibration lines, corresponding to equal phase e.m.f.s in the frames, are shown at the end of the record, and it is seen from these that the phase difference for the East-West frames is 118° , and for the North-South frames, 36° . Assuming the rotations were such that the former was negative and the latter positive, these lead at once to the values

Bearing	287°
Angle of incidence	12.2°

The geographical bearing is 295° , which is in as good agreement as would be expected for such steep rays, while the angle of incidence corresponds to reflexion from a height of 303 km., which again is approximately correct for the F region at this time of day. Other interpretations of the sense of rotation give widely different results.

(2) *Unresolved F echoes.*

During the Daventry transmissions, the two magneto-ionic components converged with increasing F density, and ultimately united into a single echo. It was of interest, therefore, to compare the behaviour of this composite echo with that of the separate components, and figure 7 shows the results of a directional test made for this purpose. There is a marked difference between the two types of record. In this case, the amplitudes of the two ellipses are seen to vary very greatly in the course of a few seconds, the fluctuations being periodic, and such that the e.m.f. in one pair of frames is weak when the other is strong. This will be recognized as a result of interference between two oppositely polarized components, and we shall refer to it further in regard to polarization; but it is particularly interesting to notice that although the ellipses are changing rapidly in size, they retain their shape fairly closely throughout the course of the run. The direction of arrival is thus constant to within about a degree, in spite of the large changes in form of the resultant wave.

This constancy of the ellipse shape is significant, also, in relation to the behaviour of the apparatus. It has been mentioned that the aeriels were designed to make the direction finder substantially free from polarization error; and it is obvious that this error would be most apparent under the conditions with which we are here dealing, when each frame is subject to a magnetic field changing widely in its direction relative to the receiving system. Although the accuracy of the test is not great, the fact that the measured direction remained practically constant in this case suggests that such errors are small, and the direction finder is reliable as far as polarization errors are concerned.

(3) *Carrier observations.*

The same recording technique was applied to a steady carrier from Daventry, for comparison with the pulse observations. The echo-selecting circuit now served merely to break the carrier up into elements of very short duration, so that separate ellipses should be produced on the film, and the duration of the active period was kept the same as before for uniformity.

An example of this type is shown in figure 8. The incident wave now comprised not only the two magneto-ionic components, but also multiple reflexions from the F layer, and the effect of their mutual interference is clearly shown. In addition to the amplitude changes of the last record, there are now large phase changes in either pair of frames, indicating a fluctuation in direction of the effective phase surface. This is as we should expect with rays arriving at different angles of incidence. Such rays add

together to produce a resultant which may vary over the range of angles embraced by the different rays, and may even extend far outside this range when the resultant amplitude is small compared with those of the individual rays. It is interesting to notice, further, that the axes of the ellipses now depart very considerably from lines at 45° . This implies that the e.m.f.s in opposite frames differ in amplitude as well as in phase, and it is seen that the e.m.f. in one frame was often as much as three times that in the other. This is attributable, again, to rays arriving at different angles of incidence. Such rays, having random phase relations to one another, may add together at one frame and oppose at the other, causing large differences in the resulting signal strengths. In the manual technique, the effect of this would be to necessitate large adjustments of the amplitude control in the phase unit, as well as of the phasing circuits, and the difficulty of following the changes in direction would be greatly increased.

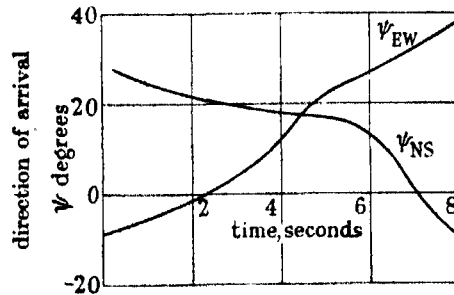


FIGURE 9. Directional variations of Daventry carrier wave, 16 May 1940.
Frequency 6.1 Mc./sec.

From the shapes of the ellipses, the directional variations may easily be computed, and figure 9 shows the results for the carrier transmission. The East-West and North-South frames are taken separately, the origin in each case being arbitrary, since there were insufficient data to determine the actual direction at any moment. It is seen that the apparent ray direction varies through as much as 30° relative to the axes of the frames in the 10 sec. of the record. This is 150 times the variation noted on the single extraordinary echo, and shows clearly the difference between carrier and pulse transmissions as regards stability of the wave surface. In direction-finding problems, it is clear from this that great advantages are to be obtained by the use of pulses as compared with ordinary continuous wave transmissions.

(4) *E reflexions.*

Using the local transmitter, observations were made on an occasion when abnormal E echoes were present. The frequency used was 6.4 Mc./sec., and the echoes were of the sporadic, scattered type, lasting only a few minutes, rather than the intense E reflexions which sometimes persist for long periods. The results of one test are shown in figure 10. The measurements were made, as before, on the 80 m. spaced frames, and the sub-multiplier was set to a frequency of $1/8$ of that of the mains.

The echoes show a remarkable variability in direction, so rapid that in places it is not even possible to follow the continuity of phase between successive ellipses. The echoes in this case were much weaker than those previously observed, and it is possible that they were influenced by noise arising from the increased gain of the receiver, which tends to introduce random changes of phase; but it is clear, where continuity exists, that large changes occur within a very few cycles of the submultiplier, and in a period of 1 sec. the change of phase is of the order of 180 – 360° . This corresponds to a directional change of 20 – 40° .

Although the possible types of E echoes differ greatly in characteristics, this result accords with the fact that they always show much poorer balances on the phase unit than echoes from the F region; and in the case of a diffuse echo, such as that recorded here, they are known to represent angular spreads of the rays of 30° or more. The two methods of measurement are thus consistent, while the present results show approximately the rate of change of direction of the effective wave surface. It would be necessary to balance the phase unit in a time of the order of $1/2$ sec. to obtain a well-defined direction by that method.

(5) *Scatter echoes.*

The short-lived scatter echoes which occur at any time on a high-power transmission have also been observed. The Ongar station, on a frequency of 7.6 Mc./sec., was used, and figure 11 shows a directional record obtained by this means. Only one transmission has been available since the completion of the apparatus, and the echo strengths were unfortunately poor on this occasion, so that noise level was again detracting from the accuracy; but the main features of the echoes may be seen. The submultiplier was put out of operation, so that the maximum repetition frequency of the ellipses should be obtained, and the film speed was increased correspondingly.

It is seen that, as in the case of abnormal E echoes, large phase changes occur within a fraction of a second. On either pair of frames there are changes of the order of 180° in $1/4$ sec., which, for the wave-length of

40 m., implies a directional change of about 15° in this time. This very high rate of change accords again with the fact that only the poorest balances can be obtained manually on these echoes, and their directional properties have, in consequence, remained largely unknown. Although we have not enough data for a detailed analysis, the results obtained throw some light on the nature of the scatter clouds which are responsible for these echoes. When observed on a time base in the usual way, each echo is seen to last for a time of the order of a few seconds, during which its effective height remains sensibly constant; and this suggests that the reflexion is due to a cloud of ionization, which spreads out until the density is insufficient to return the waves any longer. If this is correct, it would be expected that the direction of arrival of each echo should be practically constant while the cloud lasts, being the direction of the cloud itself, whereas the evidence all seems to point to very large variations in direction being present. This feature is difficult to interpret. There seems to be only one way to reconcile the two results, namely, to assume that each echo, instead of being due to a single cloud, is due to a number of such centres distributed widely in the ionosphere, such as might be produced by a shower of particles entering the atmosphere. The interference between the rays from such a group of clouds would account for the large variations in direction of the wave front, while their relative constancy of position during their lifetime would explain the apparent stability of the echo on the time base. It would be out of place here to consider the mechanism by which such a cloud structure could suddenly be brought into being, but we may state with some certainty that the echoes are of such a nature as to be inexplicable in terms of single clouds in the ionosphere, unless these are assumed to move during their lifetime with enormous and random velocities; and other considerations make this supposition very unlikely.

(b) *Polarization measurements*

(1) *Split F echoes.*

Records of polarization on the two magneto-ionic components were taken during the early morning transmissions from Daventry. The 20 m. frames were connected in parallel as described, and the submultiplier switched to give one pulse instead of two in each cycle. Figure 12 shows typical results for the two waves, the upper record being of the ordinary and the lower of the extraordinary component.

The ellipses on the film are seen to be extremely constant in shape for either component, regardless of changes in amplitude. Both are nearly circular. The angle of incidence was small, being about 12° in this experi-

ment, so that the correction for phase difference between opposite frames does not modify this result greatly. From consideration of the quadrant in which the calibration line lies, it is easy to see that the figure on the film, when corrected for this phase difference, represents the actual magnetic ellipse on the ground, where the North-South direction is taken along the length of the record. The mean results deduced by this means are:

Wave	Ratio of axes	Bearing of major axis from North
Ordinary	0.91	-35°
Extraordinary	0.92	+10°

Simultaneously with these tests, observations were made on a polarimeter in the receiving hut (Eckersley and Millington 1939). Owing to the presence of other apparatus, this was known to be subject to some local distortion, and the results obtained always differed a little from those derived by the later method. For comparison, we give the average results obtained on this apparatus, again for the ellipse on the ground:

Wave	Ratio of axes	Bearing of major axis from North
Ordinary	0.80	+46°
Extraordinary	0.71	+41°

Since the ellipses were all nearly circular, little importance can be attached to the measurement of bearing of the major axis, the position of this being vaguely defined in the field of the wave. The agreement between the two sets of results is not very good, however, and it is probable that the error resides mainly in the polarimeter values.

The polarizations to be expected in these experiments may be calculated from the well-known Appleton-Hartree theory. In our case, taking the angle of incidence to be 12°, and omitting the damping factor, the ellipse on the wave front is found to have a ratio of axes of 0.966. The value is the same for both components, their major axes being at right angles.

The ellipses on the ground are found by projecting these, using the analysis of Eckersley and Millington mentioned above. The resulting figures are then:

Wave	Ratio of axes	Bearing of major axis from North
Ordinary	0.95	+48°
Extraordinary	0.96	-14°

Both ellipses are thus within 5% of circular. It is clear, therefore, that if these values do obtain in practice, very careful measurements would be necessary to distinguish the ellipses from circles, or to determine the

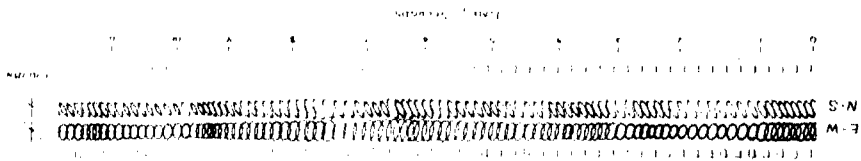


Figure 6. E extraordinary echo directions, Daventry 6.1 Mc./sec.

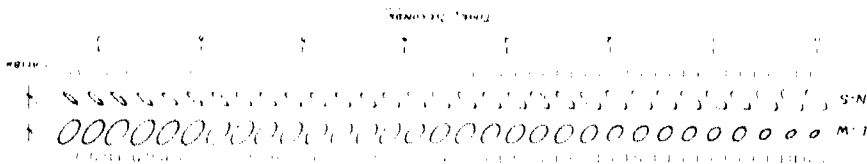


Figure 7. Unresolved F echo directions, Daventry 6.1 Mc./sec.

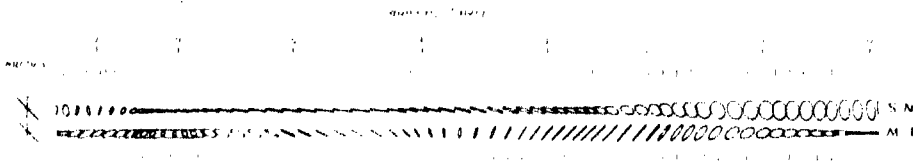


Figure 8. Daventry carrier wave directions, 6.1 Mc./sec.

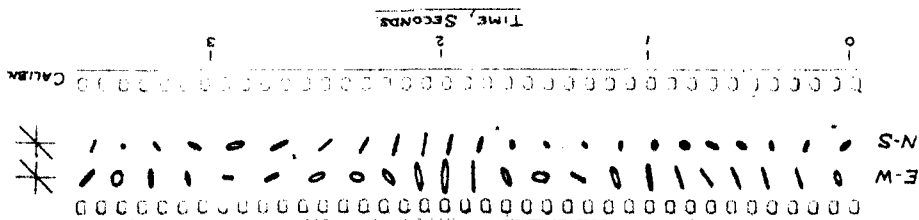


Figure 10. Diffuse J5 echo directions, 6.4 Mc./sec.

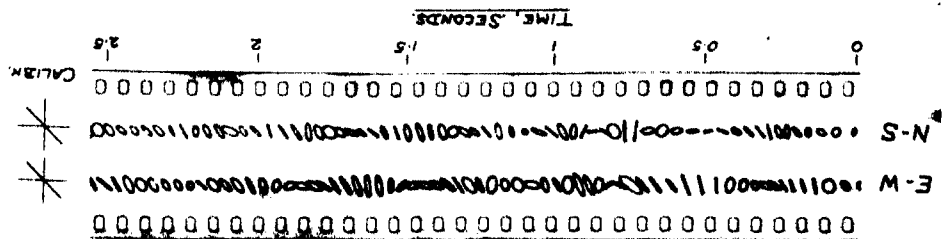


Figure 11. Scatter echo directions, Ungar 7.6 Mc./sec.

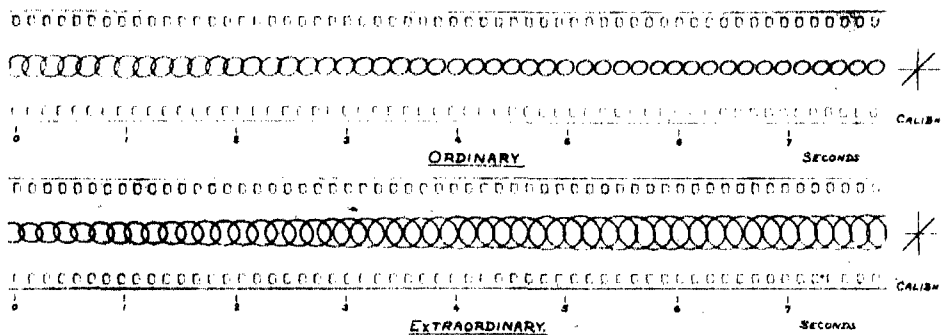


FIGURE 12. Split F echo polarizations, Daventry 6.1 Mc./sec.

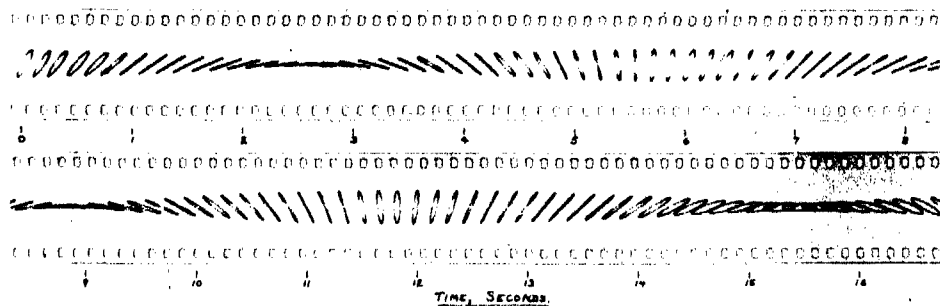


FIGURE 13. Unresolved F echo polarizations, Daventry 6.1 Mc./sec.
Ionization increasing.

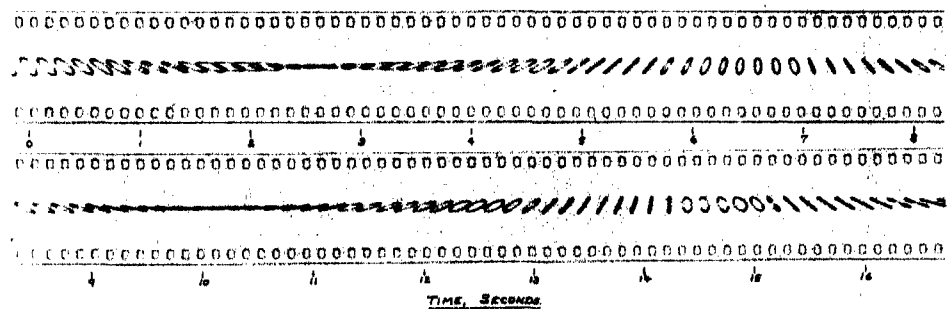


FIGURE 14. Unresolved F echo polarizations, Daventry 6.1 Mc./sec.
Ionization decreasing.

orientations of their axes. The accuracy in our experiments seems scarcely sufficient for this purpose.

There was, however, a definite tendency in the experimental results for the ellipses to be more eccentric than the theoretical figures, the ratio of axes seldom exceeding 0.9; and it seemed that this might be significant in regard to the conditions at the bottom of the layer determining the limiting polarization of the waves. It is assumed in the theory that the polarization is that characteristic of zero ionization density, and also that the damping term is negligible. That the latter supposition could not be responsible for the observed eccentricities is clear, since the inclusion of this term always makes the polarizations more circular; but it is well known that, at levels of greater density, the ellipses do become more eccentric, and it seemed possible that the form of the down-coming wave might be determined at some level slightly above the bottom of the layer, instead of at its lower boundary. From the relation between density and polarization, it is easy to calculate the shape of the ellipse corresponding to any depth in the layer, and it may be shown that to obtain a ratio of axes of 0.9 (the approximate value on the wave front from our results), it would be necessary to go to a depth where the density is 0.4 of the critical value for the ordinary wave. This is a large departure from the zero ionization at the boundary, and it is clear that if the polarization is actually determined at such a level, very considerable error is presented in the ray theory, according to which the polarization must remain characteristic throughout the wave path. The fact that this theory accords so well in other respects with the behaviour of down-coming waves, suggests that such a large departure in this case is unlikely; and it seems that a systematic error of about 5% in the measurements must be postulated to account for our observed ellipse shapes.

(2) *Unresolved F echoes.*

It is well known that when the two magneto-ionic components merge together, the polarization ceases to retain its simple form, and is determined by the interference between the two waves. The effect of this is commonly observed by the rhythmic fading it produces on an F echo just after the two components have united.

Observations were made on the F echoes from Daventry under these conditions, and figures 13 and 14 show the resulting polarization measurements. The first was taken when the two echoes had just come together; the second, just before they separated again owing to a temporary decrease in ionization density. The addition of the two components to form a plane polarized wave is clearly shown in both cases, as well as the reversal in

sense of rotation accompanying the change from increasing to decreasing density. The rate of rotation of the plane gives the speed at which one phase path was changing relative to the other, which is seen to be of the order of one wave-length in 10–20 sec.

When the echoes were more completely united, the polarization showed less definite characteristics, and remained much more steady with time. As the sun's elevation increased, it approached more and more to the form of the ordinary wave, showing that the extraordinary component was being absorbed more heavily; but there was no tendency to reach a steady elliptical state, as exhibited by one component alone.

(3) *Scatter echoes.*

No polarization records have been obtained on these echoes, but they were observed visually on the oscillograph before the photographic equipment was completed. Their characteristics may be summed up as follows:

(a) The polarization may have any elliptical form whatever, and often changes through the extreme range from a left-handed to a right-handed circle in a fraction of a second.

(b) When an echo is first produced, there is a tendency for the wave to be plane polarized, and, in the case of the Ongar station, to have its magnetic vector approximately in the East-West direction. It is possible that this is due to 'mirror' reflexion at the sharp boundary of a cloud. The transmitting aerial was a horizontal dipole lying in the direction 150° from North, and would therefore radiate a wave having its magnetic vector more in the East-West than in the North-South direction.

(c) After the echo has existed for a few seconds, it may show any of the following features: (1) The ellipse may take the form of a line rotating rapidly in either direction. This suggests, as noted above, the interference between two oppositely polarized components of approximately equal amplitudes. (2) The ellipse may retain an approximately constant shape, but fluctuate in amplitude. This would be caused by two right-hand or left-hand polarized waves of varying phase paths beating together. (3) The ellipse may oscillate within a rectangular envelope, passing through a sequence of forms as shown in figure 15. These imply a variation of phase difference between the e.m.f.s in the two pairs of frames, each one retaining an approximately constant amplitude, and would suggest the interference of two plane waves oriented at right angles at the receiving system.

In the case of a scatter echo which lasted for several minutes, it was observed that the polarization ultimately became much steadier, and tended towards a right-hand circular form. This persisted for some time, after which the amplitude decreased and the echo finally disappeared. This seems to

imply that such an echo is due to the production of exceptionally dense clouds, which diffuse to form a layer in the ionosphere; and as this diffuses further, it reflects for a time the extraordinary wave only, which also ultimately penetrates, causing loss of the echo.

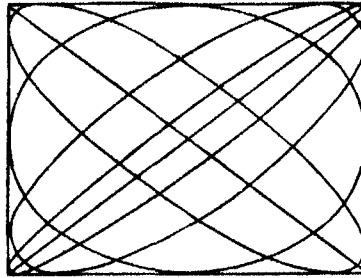


FIGURE 15. Fluctuating polarization ellipses.

It should be noted that none of the characteristics mentioned under (c) is well defined, or persists for any length of time; but their occurrence, even for short periods, throws some light on the structure of the scatter reflexions. It seems certain that these echoes are due to a large number of contributing rays, which interfere with one another in various ways; and this again points to the supposition made above that each echo is the result, not of reflexion from a single cloud, but rather from a group of clouds extending over a considerable area in the ionosphere. Further observations of the type described here will perhaps give a more complete picture of the mechanism of these reflexions.

CONCLUSIONS

The experiments as a whole throw some light on the type of irregularity that is encountered in wireless transmission in the ionosphere. From the various echoes examined, it is notable how the degree of irregularity, both of polarization and direction, varies greatly according to the type of ray received; the F reflexions, for instance, observed by the rapid technique, show scarcely any change in characteristics over a period of the order of 20 sec., and may be followed with quite a slow observing speed, while the abnormal E echoes, and more especially scatter reflexions, show far greater fluctuations, and change their characteristics considerably in a second or less. This accords, in general, with the conclusions from manual experiments; but the present observations have shown, too, the order of the upper limit that exists to the rate of change of either characteristic. It was found, throughout, that except when noise level in the amplifiers was liable

to introduce random phase changes, there was definite continuity of phase from one 50-cycle pulse to the next, even in the most rapidly changing echoes. This result is of interest, in showing that at any instant there is a perfectly definite wave surface at the receiver, which persists for a time extremely long compared with the period of oscillation of the wave. Although it may still be convenient to picture a bad balance on the phase unit as being due to a bundle of rays arriving simultaneously at the receiver, it is clear that this is only the result of the long time required to measure the phase relations between opposite frames, and that if a balance could be made instantaneously, the variations in direction could be followed in detail just as they can for the more slowly changing F echoes. In both cases, the measurement gives simply the direction of the normal to the surface of constant phase, which may, as pointed out above, be quite unrelated to the real direction of arrival of the contributing rays.

The scatter echoes show remarkable variability of both direction and polarization, and it is clear that they differ very radically from the more normal ionospheric reflexions. Several suggestions have been put forward to account for their origin. Of these, the possibility that they are due to meteors, recently proposed by Skellett (1938), deserves mention, since such a mechanism might account for a rapid movement of the scattering centre during the early part, at least, of its lifetime. An examination of the records, however, does not seem to make even this interpretation likely; for if we assume a distance away of the clouds of about 100 km., an average value for these echoes, the directional changes imply a velocity of the order of 100 km./sec. transverse to the direction of the ray. This is greatly in excess of ordinary meteor velocities. Thus it appears that any explanation in terms of a movement of the scattering source is untenable, and we are forced to the conclusion already propounded, that deviations of the wave surface are only apparent variations of the echo direction, and arise from interference between a number of different rays originating at widely spaced centres in the ionosphere.

It has been mentioned that the variability of both E and F echoes was probably due to the same cause, namely, irregularities in the E layer of the ionosphere. Whether this can be the case, in view of the enormous difference in the rates of change in the two types of reflexion, is clearly open to question, though there is reason to believe that this would be a possible consequence of such a common origin. The waves reflected back from clouds in the E region, either as scatter or diffuse E echoes, are probably due to reflexion at a sharp gradient of ionization which lasts only for a short period after the formation of the clouds, and it is clear that during

this period, conditions are changing extremely quickly, so that rapid phase changes in the reflected rays would naturally result. These would give the rapidly varying phase surface observed at the receiver. In the case of F reflexions, however, the clouds would continue to deviate for a time very long compared with that of back scattering, owing to the much smaller density required for deviation than for reflexion; and if this is the case, the bending of the waves would be caused, not so much by the clouds formed in the previous 2 or 3 sec., as by the much larger number formed within the last 10 or 15 min. The rates of change of the latter would be much slower than those of the newly formed clouds, and a relatively slow deviation would be likely to result. A mathematical analysis of this problem for the case of spherically symmetrical clouds has been given by Eckersley (1932), which shows this fundamental difference between backward and forward scattering, and supports the view outlined here. Whether the common deviating mechanism is, in fact, the correct interpretation of the phenomena, can only be decided with certainty, it seems, by a more detailed study of these scattered echoes.

In conclusion, it should be mentioned that in the present experiments only a very limited amount of data has been obtainable, and many features require further study. The technique, however, appears to be one having considerable usefulness in this direction, and it is hoped that further experiments will be possible to derive data for a more complete analysis.

We are indebted to Mr L. J. van Rooyen for the design and construction of much of the apparatus used in this work, and to Mr S. W. H. W. Falloon for valuable suggestions concerning the technique, and to Marconi's Wireless Telegraph Co., Ltd., for permission to publish these results. It is hoped to publish an account of the experimental equipment separately.

REFERENCES

- Barfield, R. H. and Ross, W. 1938 *J. Inst. Elect. Engrs*, **83**, 98.
Eckersley, T. L. 1932 *J. Inst. Elect. Engrs*, **71**, 405.
— 1935 *Marconi Rev.* **53**, 1.
— 1938 *Nature, Lond.*, **141**, 369.
— 1940 *J. Inst. Elect. Engrs*, **86**, 548.
Eckersley, T. L. and Millington, G. 1939 *Proc. Phys. Soc.* **51**, 110.
Farmer, F. T. 1935 *Proc. Camb. Phil. Soc.* **31**, 295.
Martyn, D. F. and Green, A. L. 1935 *Proc. Roy. Soc. A*, **148**, 104.
Martyn, D. F., Piddington, J. H. and Munro, G. H. 1937 *Proc. Roy. Soc. A*, **158**, 536.
Ratcliffe, J. A. and Pawsey, J. L. 1933 *Proc. Camb. Phil. Soc.* **29**, 301.
Skelllett, A. M. 1938 *Nature, Lond.*, **141**, 472.

Complex potentials in two-dimensional elasticity. II

By A. C. STEVENSON

(Communicated by G. B. Jeffery, F.R.S.—Received 11 December 1940)

The complex potentials are found for two elastic problems, involving body forces, for a circular disk.

INTRODUCTION

This paper gives the exact solution in finite terms for the complex potentials which solve two problems of generalized plane stress for a circular disk. The first of these concerns the two-dimensional elastic problem of a heavy circular disk fixed in a vertical plane at an eccentric point, and so includes as particular cases that for which the point of fixation is on the boundary of the disk, previously solved by the writer (Stevenson 1945, p. 176), and also that for which the disk is merely suspended freely from an eccentric point, the Airy stress function for which was recently obtained by Mindlin (1938, p. 714) employing solutions of the biharmonic equation in bipolar co-ordinates (Jeffery 1921, p. 265). It will be shown that the complex potentials consist of combinations of a few simple terms in cartesian (complex) co-ordinates. The second problem concerns the rotation of a circular disk in its own plane at a steady rate about an eccentric point. Here again the complex potentials consist of a few simple cartesian terms.

1. STRESSES IN A HEAVY CIRCULAR DISK FIXED AT ONE POINT

We shall take the origin $z = 0$ at the centre O of the disk, so that the equation of the boundary in complex co-ordinates is $z\bar{z} = a^2$. The point of fixation C will be taken as $z = -if$, where f is real and less than a . The disk will be taken as of unit thickness, and we consider the elastic problem as one of generalized plane stress, the plane of the disk being vertical. If the direction of gravity makes an angle α with the x -axis, the displacements and stresses are given in terms of complex potentials $\Omega(z)$, $\omega(z)$ by the results (Stevenson 1945, pp. 133, 138)

$$8\mu(u + iv) = \kappa\Omega(z) - z\bar{\Omega}'(\bar{z}) - \bar{\omega}'(\bar{z}) + 4\rho\left(\frac{\kappa - 1}{\kappa + 1}\right)\left(\frac{1}{2}cz^2 + \bar{c}z\bar{z}\right), \quad (1.1)$$

$$\bar{x}\bar{x} + \bar{y}\bar{y} = \frac{1}{2} \left\{ \Omega'(z) + \bar{\Omega}'(\bar{z}) + \frac{8\rho}{\kappa+1} (cz + \bar{c}\bar{z}) \right\}, \quad (1.2)$$

$$\bar{x}\bar{x} - \bar{y}\bar{y} + 2i\bar{x}\bar{y} = -\frac{1}{2} \left\{ z\bar{\Omega}''(\bar{z}) + \bar{z}\Omega''(z) - 4\rho \left(\frac{\kappa-1}{\kappa+1} \right) \bar{c}z \right\}, \quad (1.3)$$

$$\text{where} \quad c = -\frac{1}{2}ge^{-i\alpha}, \text{ and } (1+\kappa)(1+\eta) = 4, \quad (1.4)$$

η being Poisson's ratio and ρ the density of the elastic material.

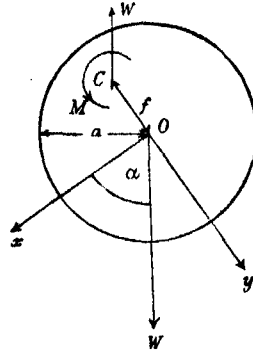


FIGURE 1. Forces on the disk.

The body forces are equivalent to the weight $W = \pi\rho ga^2$ localized at the centre of mass O of the disk and acting at an angle α with the x -axis. At the nucleus of fixation C therefore, the stresses round a small circuit about C are equivalent to a force X, Y at C , together with a couple M , given by

$$X + iY = \pi\rho ga^2 e^{i\alpha}, \quad M = -\pi\rho ga^2 f \cos \alpha. \quad (1.5)$$

The complex potentials appropriate to the origin C for this force and couple nucleus are known to be (Stevenson 1945, p. 154)

$$\Omega_C(z_1) = A \log z_1, \quad \omega_C(z_1) = -\kappa \bar{A} z_1 \log z_1 + B \log z_1,$$

where $z_1 = z + if$, and A and B are constants given by

$$\left. \begin{aligned} A &= 2\rho ga^2 e^{i\alpha} / (\kappa + 1) = -4\rho a^2 \bar{c} / (\kappa + 1), \\ B &= 2i\rho ga^2 f \cos \alpha = \frac{1}{2} if (\kappa + 1) (A + \bar{A}). \end{aligned} \right\} \quad (1.6)$$

If $z = re^{i\theta}$, the stresses appropriate to circular boundaries $r = \text{const.}$ can be expressed in complex form (Stevenson 1945, p. 146) as

$$4(\bar{r}\bar{r} - ir\bar{\theta}) = \Omega'(z) + \bar{\Omega}'(\bar{z}) - z\bar{\Omega}''(\bar{z}) - \bar{z}\Omega''(z) - \frac{z}{\bar{z}}\omega''(\bar{z}) + 4\rho cz + \frac{8\rho\bar{c}\bar{z}}{\kappa+1}. \quad (1.7)$$

For the origin O , the complex potentials $\Omega(z), \omega(z)$ are given (Stevenson 1945, p. 143) as

$$\Omega(z) = \Omega_C(z_1), \quad \omega(z) = \omega_C(z_1) - if\Omega_C(z_1),$$

or, using (1.6)

$$\Omega(z) = A \log z_1, \quad \omega(z) = -\kappa \bar{A} z_1 \log z_1 + \frac{1}{2} if \{(\kappa-1)A + (\kappa+1)\bar{A}\} \log z_1. \quad (1.8)$$

From (1.7), using (1.6) and (1.8),

$$\begin{aligned} 4(\widehat{rr} - i\widehat{r}\widehat{\theta}) &= \frac{A}{z_1} + \frac{\bar{A}}{\bar{z}_1} + \frac{Az}{z_1^2} + \kappa \frac{\bar{A}z}{\bar{z}_1 z_1} + \frac{1}{2} if \{(\kappa-1)A + (\kappa+1)\bar{A}\} \frac{z}{\bar{z} z_1^2} \\ &\quad - (\kappa+1) \frac{\bar{A}}{a^2} z - 2 \frac{A}{a^2} \bar{z}. \end{aligned} \quad (1.9)$$

Now on the boundary $z\bar{z} = a^2$, we have

$$z_1 = z + if = if + \frac{a^2}{z} = \frac{if}{z} \left(\bar{z} + \frac{a^2}{if} \right)$$

$$\text{or} \quad z_1 \bar{z} = if \bar{z}_2, \quad \text{where} \quad z_2 = z + i \frac{a^2}{f}. \quad (1.10)$$

We now express all the terms in the right-hand side of (1.9) in terms of z_2, \bar{z}_2 . Thus we have

$$\frac{1}{z_1} = \frac{\bar{z}}{if \bar{z}_2} = \frac{\bar{z}_2 + ia^2/f}{if \bar{z}_2} = -\frac{i}{f} + \frac{a^2}{f^2} \frac{1}{\bar{z}_2},$$

$$\frac{z}{\bar{z} z_1} = \frac{a^2}{\bar{z}} \frac{1}{if \bar{z}_2} = \frac{1}{\bar{z}} - \frac{1}{\bar{z}_2} = \frac{z_2}{a^2} - \frac{i}{f} - \frac{1}{\bar{z}_2},$$

$$\frac{z}{\bar{z} z_1^2} = \frac{z \bar{z}}{(\bar{z} z_1)^2} = -\frac{a^2}{f^2} \frac{1}{\bar{z}_2^2},$$

$$\text{and} \quad \frac{z}{z_1^2} = \bar{z} \left(\frac{z}{\bar{z} z_1^2} \right) = -\frac{a^2}{f^2} \frac{1}{\bar{z}_2} - i \frac{a^4}{f^2} \frac{1}{\bar{z}_2^2}.$$

From these results, (1.9) becomes, along the circular boundary of the disk,

$$\begin{aligned} 4(\widehat{rr} - i\widehat{r}\widehat{\theta}) &= \frac{i}{f} (2\bar{A} - 3A) - \kappa \frac{\bar{A}}{\bar{z}_2} - 2 \frac{A}{a^2} \bar{z}_2 - \frac{\bar{A}}{a^2} z_2 + \bar{A} \frac{a^2}{f^2} \frac{1}{z_2} \\ &\quad - \frac{ia^2}{2f \bar{z}_2^2} \left\{ 2 \frac{a^2}{f^2} A + (\kappa-1)A + (\kappa+1)\bar{A} \right\}. \end{aligned} \quad (1.11)$$

Consider now the complementary complex potentials

$$\Omega(z) = Cz^2 + Dz + \frac{E}{z_2} + F \log z_2, \quad \omega(z) = Gz_2 \log z_2 + K \log z_2 + \frac{L}{z_2}, \quad (1.12)$$

where C, \dots, L are complex constants, of which D may be taken real, since its imaginary part corresponds merely to a rigid body displacement. Since the point $z = z_2$ lies outside the circular boundary of the disk, these complex potentials clearly yield physically admissible stresses and displacements (Stevenson 1945, p. 139). From (1.7), with $c = 0$, the corresponding stresses appropriate to a boundary $r = \text{const.}$ are given by

$$4(\bar{r}r - i\bar{\theta}\theta) = 2D + 2\bar{C}\bar{z} - \frac{E}{z_2^2} - \frac{\bar{E}}{\bar{z}_2^2} + \frac{F}{z_2} + \frac{\bar{F}}{\bar{z}_2} - z\left(2\frac{E}{z_2^3} - \frac{F}{z_2^2}\right) - \bar{z}\left(\frac{G}{z_2} - \frac{K}{z_2^2} + 2\frac{L}{z_2^3}\right),$$

which can be written, along the boundary $z\bar{z} = a^2$, in terms of z_2, \bar{z}_2 , since $z = z_2 - ia^2/f$, as

$$\begin{aligned} 4(\bar{r}r - i\bar{\theta}\theta) = & \left\{ 2D + 2i\bar{C}\frac{a^2}{f} + \frac{K}{a^2} + 2i\frac{G}{f} \right\} + 2\bar{C}\bar{z}_2 + \frac{\bar{F}}{\bar{z}_2} - \frac{\bar{E}}{\bar{z}_2^2} - \frac{G}{a^2}z_2 \\ & + \frac{1}{z_2} \left\{ 2F - 2\frac{L}{a^2} - 2i\frac{K}{f} + G\frac{a^2}{f^2} \right\} \\ & + \frac{1}{z_2^2} \left\{ 4i\frac{L}{f} - 3E - i\frac{a^2}{f}F - K\frac{a^2}{f^2} \right\} + 2\frac{a^2}{f^2}(L + ifE)\frac{1}{z_2^2}. \quad (1.13) \end{aligned}$$

Combining the results of (1.11) and (1.13), we see that we have an unstressed boundary if we can choose the coefficients C, \dots, L so that the coefficients of the various different terms vanish simultaneously. This necessitates

$$\left. \begin{aligned} \bar{F} - \kappa\bar{A} &= 0, \quad \bar{C} - A/a^2 = 0, \quad G + \bar{A} = 0, \quad L + ifE = 0, \\ -\bar{E} - i\frac{a^2}{2f} \left\{ 2\frac{a^2}{f^2}A + (\kappa-1)A + (\kappa+1)\bar{A} \right\} &= 0, \quad 4i\frac{L}{f} - 3E - i\frac{a^2}{f}F - K\frac{a^2}{f^2} = 0, \end{aligned} \right\} \quad (1.14)$$

$$2F - 2\frac{L}{a^2} - 2i\frac{K}{f} + G\frac{a^2}{f^2} + \bar{A}\frac{a^2}{f^2} = 0, \quad (1.15)$$

$$2D + 2i\bar{C}\frac{a^2}{f^2} + \frac{K}{a^2} + 2i\frac{G}{f} + 2i\frac{\bar{A}}{f} - 3i\frac{A}{f} = 0. \quad (1.16)$$

The six equations (1.14) give the six constants F, C, G, E, L, K in terms of A as

$$\left. \begin{aligned} F &= \kappa A, \quad C = \bar{A}/a^2, \quad G = -\bar{A}, \\ E &= i\frac{L}{f} = i\frac{a^2}{2f} \left\{ 2\frac{a^2}{f^2}\bar{A} + (\kappa-1)\bar{A} + (\kappa+1)A \right\}, \\ K &= \frac{1}{2}if \left\{ 2\frac{a^2}{f^2}\bar{A} + (\kappa-1)(\bar{A} - A) \right\}, \end{aligned} \right\} \quad (1.17)$$

and these values satisfy (1.15), whilst the remaining equation (1.16) gives for the constant D

$$D = \frac{1}{2}i \frac{f}{a^2} \left(\kappa + 2 \frac{a^2}{f^2} - 1 \right) (A - \bar{A}), \quad (1.18)$$

which is satisfactorily real as assumed. Combining the complex potentials (1.8) and (1.12), using the results (1.4), (1.6), (1.17) and (1.18), we see that the complex potentials which solve the problem of the heavy circular disk, fixed at one point, are given by

$$\Omega(z) = \frac{2\rho g}{\kappa+1} \left\{ a^2 e^{i\alpha} \log z_1 + \kappa a^2 e^{i\alpha} \log z_2 + e^{-i\alpha} z^2 - \frac{1}{2} f \sin \alpha \left(\kappa + 2 \frac{a^2}{f^2} - 1 \right) z \right. \\ \left. + \frac{a^4}{f} \left[\left(\frac{a^2}{f^2} - 1 \right) \sin \alpha + i \left(\kappa + \frac{a^2}{f^2} \right) \cos \alpha \right] \frac{1}{z_2} \right\}, \quad (1.19)$$

$$\omega(z) = \frac{2\rho g a^2}{\kappa+1} \left\{ -\kappa e^{-i\alpha} z_1 \log z_1 - e^{-i\alpha} z_2 \log z_2 + f (\sin \alpha + i \kappa \cos \alpha) \log z_1 \right. \\ \left. + f \left[\left(\kappa + \frac{a^2}{f^2} - 1 \right) \sin \alpha + i \frac{a^2}{f^2} \cos \alpha \right] \log z_2 \right. \\ \left. + a^2 \left[\left(\kappa + \frac{a^2}{f^2} \right) \cos \alpha + i \left(1 - \frac{a^2}{f^2} \right) \sin \alpha \right] \frac{1}{z_2} \right\}. \quad (1.20)$$

If we put $\alpha = \frac{1}{2}\pi$, we get the solution for the complex potentials for the freely hanging disk as

$$\Omega(z) = \frac{2\rho g}{\kappa+1} \left\{ i a^2 \log z_1 + i \kappa a^2 \log z_2 - i z^2 - \frac{1}{2} f \left(\kappa + 2 \frac{a^2}{f^2} - 1 \right) z + \frac{a^4}{f} \left(\frac{a^2}{f^2} - 1 \right) \frac{1}{z_2} \right\}, \quad (1.21)$$

$$\omega(z) = \frac{2\rho g a^2}{\kappa+1} \left\{ i \kappa z_1 \log z_1 + i z_2 \log z_2 + f \log z_1 + f \left(\kappa + \frac{a^2}{f^2} - 1 \right) \log z_2 + i \frac{a^2}{z_2} \right\}. \quad (1.22)$$

Mindlin (1938, p. 714) has found the Airy stress function for this particular problem in bipolar co-ordinates. This is a complicated expression in marked contrast to the relative simplicity of the results (1.21), (1.22). We shall content ourselves by comparing the expressions found by each method for the peripheral stress $\theta\theta_{r=a}$, which Mindlin has converted to polar co-ordinates for the origin O from the bipolar results. From (1.2), using (1.21), we find

$$\bar{r}r + \theta\theta = \bar{x}x + \bar{y}y = \text{real part of } \Omega'(z) + 4\rho g i z / (\kappa + 1) \\ = \text{real part of } \frac{2\rho g}{\kappa+1} \left\{ i a^2 \left(\frac{1}{z_1} + \frac{\kappa}{z_2} \right) - \frac{1}{2} \left(\kappa + 2 \frac{a^2}{f^2} - 1 \right) - \frac{a^4}{f} \left(\frac{a^2}{f^2} - 1 \right) \frac{1}{z_2^2} \right\},$$

whence, if $|z_1| = r_1$, we find

$$\partial\theta_{r=a} = \frac{2\rho g(a^2 - f^2)}{(\kappa + 1)r_1^2} \left\{ \frac{1}{2}f(\kappa - 1) - \frac{a^2}{r_1^2} \left[2f + a \left(1 + \frac{f^2}{a^2} \right) \sin \theta \right] \right\},$$

or, substituting for κ in terms of Poisson's ratio η from (1.4) and putting $f/a = d$, we obtain

$$\partial\theta_{r=a} = \frac{1}{2} \frac{\rho g a (1 - d^2)}{1 + 2d \sin \theta + d^2} \left\{ (1 - \eta)d - (1 + \eta) \frac{(1 + d^2) \sin \theta + 2d}{1 + 2d \sin \theta + d^2} \right\}. \quad (1.23)$$

This is Mindlin's result for the peripheral stress, after allowing for a difference of $\frac{1}{2}\pi$ in choosing the initial line of the polar co-ordinates.

2. CHANGE OF ORIGIN FOR COMPLEX POTENTIALS WITH GRAVITATIONAL BODY FORCE

Returning to equations (1.19), (1.20), if we put $f = -a$, we have $z_1 = z_2$ and

$$\Omega(z) = \rho g \left\{ 2a^2 e^{i\alpha} \log z_1 + \frac{2e^{-i\alpha}}{\kappa + 1} z^2 + a \sin \alpha z - 2ia^3 \cos \alpha \frac{1}{z_1} \right\}, \quad (2.1)$$

$$\omega(z) = \rho g \left\{ -2ia^3 e^{-i\alpha} \log z_1 - 2a^2 e^{-i\alpha} z_1 \log z_1 + 2a^4 \cos \alpha \frac{1}{z_1} \right\}, \quad (2.2)$$

as the solution for the problem previously considered by the writer for the case where the point of fixation C is on the edge of the disk (Stevenson 1945, p. 176). There the solution is given by our equations (1.1)–(1.4) and complex potentials all referred to the origin C . One mode of comparison of the two solutions would be to write down the solution for $8\mu(u + iv)$ in each case, change the origin in one case, and show that the results only differ by a rigid body displacement at most. A neater mode of comparison is to find the complex potentials appropriate to the origin C for those from the origin O . The rules for this change of origin in the case of no body force have been given previously (Stevenson 1945, p. 143), namely, that if $\Omega_C(z_1), \omega_C(z_1)$ are the complex potentials appropriate to the origin $C(z = z_0, z_1 = 0)$, then

$$\Omega_C(z_1) = \Omega(z), \quad \omega_C(z_1) = \omega(z) + \bar{z}_0 \Omega(z). \quad (2.3)$$

These were used in obtaining the partial complex potentials (1.8), since partial or complementary complex potentials are derived from the equations with no body forces (Stevenson 1945, p. 134). Now, keeping the terms in

c, \bar{c} in (1.1)–(1.3) invariant in form for a change of origin, and noting that the stresses must be independent of change of origin, we have from (1.2)

$$\Omega'(z) + \bar{\Omega}'(\bar{z}) + \frac{8\rho}{\kappa+1}(cz + \bar{c}\bar{z}) = \Omega'_C(z_1) + \bar{\Omega}'_C(\bar{z}_1) + \frac{8\rho}{\kappa+1}(cz_1 + \bar{c}\bar{z}_1),$$

which is satisfied by taking

$$\Omega_C(z_1) = \Omega(z) + 8\rho cz_0 z / (\kappa + 1). \quad (2.4)$$

Also from (1.3)

$$\bar{z}\Omega''(z) + \omega''(z) - 4\rho\left(\frac{\kappa-1}{\kappa+1}\right)c\bar{z} = \bar{z}_1\Omega''_C(z_1) + \omega''_C(z_1) - 4\rho\left(\frac{\kappa-1}{\kappa+1}\right)c\bar{z}_1,$$

which is satisfied by (2.4) and

$$\omega_C(z_1) = \omega(z) + \bar{z}_0\Omega(z) - 2\rho c\left(\frac{\kappa-1}{\kappa+1}\right)\bar{z}_0 z^2, \quad (2.5)$$

in which latter formula we may replace $\Omega(z)$ by $\Omega_C(z_1)$ and z^2 in the last term by z_1^2 without altering the stresses or the displacements by more than a rigid body displacement. If (2.4) and (2.5) are substituted into the right-hand side of (1.1), it will be found that

$$u + iv = u_1 + iv_1 + \alpha + i\beta z$$

(α complex, β real), so that the solutions for $\Omega_C(z_1)$, $\omega_C(z_1)$ and $\Omega(z)$, $\omega(z)$ differ only by a rigid body displacement.

Applying this method to (2.1) and (2.2), changing the origin to $z = z_0 = ia$, (2.4) and (2.5) lead to

$$\Omega_C(z_1) = \rho g \left\{ 2a^2 e^{i\alpha} \log z_1 + \frac{2e^{-i\alpha}}{\kappa+1} z_1^2 + a \sin \alpha z_1 - 2ia^3 \cos \alpha \frac{1}{z_1} \right\}, \quad (2.6)$$

$$\omega_C(z_1) = -\rho g \{ 4ia^3 \cos \alpha \log z_1 + 2a^3 e^{i\alpha} z_1 \log z_1 + iae^{-i\alpha} z_1^2 \}, \quad (2.7)$$

where an irrelevant constant has been added to the right-hand side of (2.6). These are the results previously found by the writer for this particular case (Stevenson 1945, p. 177).

3. STRESSES IN A DISK ROTATING ABOUT AN ECCENTRIC POINT

We shall suppose a light circular disk of centre O and radius a (see figure 1) is rotated in its plane about a smooth axis through an eccentric point $C(z = -if)$ with a constant angular velocity n . The problem is considered as a statical problem, the reversed mass-accelerations being treated

as body forces in comparison with which the gravitational body forces are neglected. Considering the problem as one of generalized plane stress, the solution in terms of complex potentials, with the same origin and axes at O as used in the last example, is given in polar co-ordinates (Stevenson 1945, pp. 146, 133) by

$$8\mu(u_r + iu_\theta) = e^{-i\theta} \{ \kappa \Omega(z) - z \bar{\Omega}'(\bar{z}) - \bar{\omega}'(\bar{z}) + \frac{1}{2} \gamma \rho (cz^2 + 2c\bar{z}z + d\bar{z}^2) \}, \quad (3.1)$$

$$\bar{r}\bar{r} + \bar{\theta}\bar{\theta} = \frac{1}{2} \left\{ \Omega'(z) + \bar{\Omega}'(\bar{z}) + \frac{8\rho}{\kappa+1} (cz + \bar{c}\bar{z} + d\bar{z}\bar{z}) \right\}, \quad (3.2)$$

$$\bar{r}\bar{r} - \bar{\theta}\bar{\theta} + 2i\bar{r}\bar{\theta} = -\frac{1}{2} \left\{ z\bar{\Omega}''(\bar{z}) + \frac{z}{\bar{z}}\bar{\omega}''(\bar{z}) - \gamma\rho(\bar{c}\bar{z} + \frac{1}{2}d\bar{z}\bar{z}) \right\}, \quad (3.3)$$

where $\gamma = 4(\kappa-1)/(\kappa+1)$ and $(1+\kappa)(1+\eta) = 4$, (3.4)

η being Poisson's ratio, whilst c and d are constants given by

$$c = \frac{1}{2}n^2\bar{z}_0, \quad d = -\frac{1}{2}n^2. \quad (3.5)$$

The boundary stresses appropriate to the circles $|z| = \text{const.}$ are given in complex form, from (3.2) and (3.3), following a change of sign of i , as

$$\begin{aligned} 4(\bar{r}\bar{r} - i\bar{r}\bar{\theta}) = & \Omega'(z) + \bar{\Omega}'(\bar{z}) - z\bar{\Omega}''(\bar{z}) - \frac{z}{\bar{z}}\bar{\omega}''(\bar{z}) \\ & + 2i\rho f n^2 z - \frac{4i\rho f n^2 \bar{z}}{\kappa+1} - \frac{\kappa+3}{\kappa+1} \rho n^2 \bar{z}\bar{z}. \end{aligned} \quad (3.6)$$

The reaction of the smooth axis on the material of the disk (assumed of unit thickness) is of amount $\pi\rho f a^2 n^2$ and is along OC ; hence the stresses around a small circuit about C are equivalent to a single force $\pi\rho f a^2 n^2$ along CO , i.e. to a force X , Y and couple M , where

$$X + iY = i\pi\rho f a^2 n^2, \quad M = 0.$$

To such a force nucleus at C ($z = z_0 = -if$, $z_1 = 0$) correspond complex potentials $\Omega_C(z_1), \omega_C(z_1)$ for the origin C , where (Stevenson 1945, p. 154)

$$\Omega_C(z_1) = F \log z_1, \quad \omega_C(z_1) = -\kappa \bar{F} z_1 \log z_1 = \kappa F z_1 \log z_1, \quad (3.7)$$

and $z_1 = z + if, \quad F = 2i\rho f a^2 n^2 / (\kappa + 1). \quad (3.8)$

From (2.3) the complex potentials appropriate to the origin O are then

$$\Omega(z) = F \log z_1, \quad \omega(z) = \kappa F z_1 \log z_1 - F if \log z_1. \quad (3.9)$$

From (3.8) and (3.9), (3.6) becomes

$$4(\widehat{r\bar{r}} - i\bar{r}\widehat{\theta}) = \frac{F}{z_1} - \frac{F}{\bar{z}_1} + \frac{z}{z_1^2}F - \frac{z}{\bar{z}\bar{z}_1}\kappa F - \frac{z}{\bar{z}\bar{z}_1^2}ifF + (\kappa+1)\frac{F}{a^2}z - 2\frac{F}{a^2}\bar{z} + i(\kappa+3)z\bar{z}\frac{F}{2a^2f}. \quad (3.10)$$

Along the boundary $z\bar{z} = a^2$, this can be written in terms of z_2, \bar{z}_2 , where $z_2 = z + ia^2/f$, by the same method as used in dealing with (1.9), and becomes

$$4(\widehat{r\bar{r}} - i\bar{r}\widehat{\theta}) = i(\kappa-7)\frac{1}{2f} - \frac{a^2}{f^2}\frac{1}{z_2} + \frac{\kappa}{\bar{z}_2} + \frac{z_2}{a^2} - \frac{2}{a^2}\bar{z}_2 + i\frac{a^2}{f}\left(1 - \frac{a^2}{f^2}\right)\frac{1}{\bar{z}_2^2}. \quad (3.11)$$

Now take complementary potentials, for the origin O , given by

$$\Omega(z) = A \log z_2 + Bz + Cz^2 + D/z_2, \quad \omega(z) = Gz_2 \log z_2 + K \log z_2 + L/z_2, \quad (3.12)$$

where A, \dots, L are complex constants, of which B will be taken to be real, since the imaginary part only leads to rigid body displacements. These complex potentials lead to stresses $\widehat{r\bar{r}}, \widehat{\theta\bar{\theta}}$, given from (3.2) and (3.3), in which c and d have been put equal to zero, as

$$4(\widehat{r\bar{r}} - i\bar{r}\widehat{\theta}) = \left\{ 2B + 2i\frac{a^2}{f}\bar{C} + 2i\frac{G}{f} + \frac{K}{a^2} \right\} + 2\bar{C}\bar{z}_2 - \frac{G}{a^2}z_2 + \frac{1}{z_2} \left\{ 2A + \frac{a^2}{f^2}G - 2i\frac{K}{f} - 2\frac{L}{a^2} \right\} + \frac{\bar{A}}{\bar{z}_2} - \frac{\bar{D}}{\bar{z}_2^2} + \frac{1}{z_2^2} \left\{ 4i\frac{L}{f} - iA\frac{a^2}{f} - 3D - K\frac{a^2}{f^2} \right\} + \frac{2a^2}{z_2^2 f^2} \{L + ifD\}. \quad (3.13)$$

Combining the complex potentials (3.9) and (3.12), the combined boundary stresses (3.11) and (3.13) vanish if

$$\left. \begin{aligned} \bar{C} - \frac{F}{a^2} &= 0, \quad F - G = 0, \quad \kappa F + \bar{A} = 0, \\ \bar{D} - i\frac{a^2}{f}\left(1 - \frac{a^2}{f^2}\right)F &= 0, \quad L + ifD = 0, \\ 4i\frac{L}{f} - i\frac{a^2}{f}A - 3D - K\frac{a^2}{f^2} &= 0, \end{aligned} \right\} \quad (3.14)$$

$$2A + \frac{a^2}{f^2}G - 2i\frac{K}{f} - 2\frac{L}{a^2} - \frac{a^2}{f^2}F = 0, \quad (3.15)$$

$$(\kappa-7)i\frac{F}{2f} + 2B + 2i\frac{a^2}{f}\bar{C} + 2i\frac{G}{f} + \frac{K}{a^2} = 0. \quad (3.16)$$

The six equations (3.14) then give the six constants A, C, G, D, L, K in terms of F as

$$\left. \begin{aligned} A &= \kappa F, \quad C = -\frac{F}{a^2}, \quad G = F, \quad D = i\frac{a^2}{f}\left(1 - \frac{a^2}{f^2}\right)F, \\ L &= a^2\left(1 - \frac{a^2}{f^2}\right)F, \quad K = -ifF\left(\kappa + \frac{a^2}{f^2} - 1\right). \end{aligned} \right\} \quad (3.17)$$

These values are consistent with (3.15), and (3.16) gives the remaining constant B as

$$B = -\frac{1}{2}(\kappa - 1)\left(1 - 2\frac{f^2}{a^2}\right)\frac{iF}{f}, \quad (3.18)$$

which is satisfactorily real, as assumed, using (3.8).

Hence the complex potentials which solve the problem of the circular disk rotating steadily about an eccentric point are

$$\Omega(z) = \frac{2\rho fn^2}{\kappa + 1} \left\{ ia^2 \log z_1 + \kappa ia^2 \log z_2 + \frac{1}{2}(\kappa - 1) \left(1 - 2\frac{f^2}{a^2}\right) \frac{a^2}{f} z - iz^2 \right. \\ \left. + \frac{a^4}{f} \left(1 - \frac{a^2}{f^2}\right) \frac{1}{z_2} \right\}, \quad (3.19)$$

$$\omega(z) = \frac{2\rho fn^2}{\kappa + 1} \left\{ ia^2 \kappa z_1 \log z_1 + fa^2 \log z_1 + ia^2 z_2 \log z_2 \right. \\ \left. + fa^2 \left(\kappa + \frac{a^2}{f^2} - 1\right) \log z_2 + ia^4 \left(1 - \frac{a^2}{f^2}\right) \frac{1}{z_2} \right\}. \quad (3.20)$$

These results, in conjunction with the writer's previous results for the particular case $f = 0$ disclose an interesting phenomenon of possible discontinuity in the functional forms of the complex potentials $\Omega(z), \omega(z)$. Considered as a family of solutions obtained from (3.19) and (3.20) by varying the parameter f , we might expect to find the solution for the case where the disk rotates steadily about its centre merely by allowing f to tend to zero in (3.19), (3.20). If we do this, however, it is at first disconcerting to find that the terms in $1/z_2$ become meaningless, neither do the remaining terms agree with the known solution for this case, which is (Stevenson 1945, p. 150)

$$\Omega(z) = \frac{1}{2} \frac{\kappa + 3}{\kappa + 1} \rho a^2 n^2 z, \quad \omega(z) = 0. \quad (3.21)$$

All that we can demand of the complex potentials, however, is that the two solutions must give continuity of the stresses and displacements as $f \rightarrow 0$,

and it is not difficult to show that this is the case. Thus, for example, we find from (3.2)

$$\bar{r}\bar{r} + \bar{\theta}\bar{\theta} = \frac{i\rho f n^2}{\kappa + 1} \left\{ a^2 \left(\frac{1}{z_1} - \frac{1}{\bar{z}_1} \right) + \kappa a^2 \left(\frac{1}{z_2} - \frac{1}{\bar{z}_2} \right) - \frac{ia^2}{2f} (\kappa - 1) \left(1 - 2 \frac{f^2}{a^2} \right) - \frac{ia^4}{f} \left(1 - \frac{a^2}{f^2} \right) \left(\frac{1}{z_2^2} + \frac{1}{\bar{z}_2^2} \right) + 2i \frac{z\bar{z}}{f} \right\}, \quad (3.22)$$

and as $f \rightarrow 0$, $z_2 \rightarrow \infty$, $fz_2 \rightarrow ia^2$ and

$$\bar{r}\bar{r} + \bar{\theta}\bar{\theta} \rightarrow \frac{\rho a^2 n^2}{\kappa + 1} \left\{ \frac{1}{2} (\kappa - 1) + 2 - 2 \frac{z\bar{z}}{a^2} \right\},$$

$$\text{or} \quad \bar{r}\bar{r} + \bar{\theta}\bar{\theta} = \frac{\rho n^2}{2(\kappa + 1)} \{ (\kappa + 3) a^2 - 4r^2 \}. \quad (3.23)$$

Again from (3.2) and (3.21)

$$\bar{r}\bar{r} + \bar{\theta}\bar{\theta} = \frac{\rho n^2}{2(\kappa + 1)} \{ (\kappa + 3) a^2 - 4r^2 \}, \quad (3.24)$$

in agreement with (3.23) and both give the well-known value (see for example Timoshenko 1934, p. 68) for the peripheral stress $\bar{\theta}\bar{\theta}_{r=a}$, namely

$$\bar{\theta}\bar{\theta}_0 = \bar{\theta}\bar{\theta}_{r=a} = \frac{1}{2} \rho a^2 n^2 \left(\frac{\kappa - 1}{\kappa + 1} \right) = \frac{1}{2} \rho (1 - \eta) a^2 n^2, \quad (3.25)$$

using (3.5).

In the general case, (3.22) gives for the peripheral stress $\bar{\theta}\bar{\theta}$ at $r = a$, writing $z_1 = r_1 e^{i\theta_1}$, $z_2 = r_2 e^{i\theta_2}$, whence $r_2/a = r_1/f$ on the boundary

$$\bar{\theta}\bar{\theta} = \rho n^2 \left(\frac{\kappa - 1}{\kappa + 1} \right) \left\{ \frac{1}{2} a^2 + f^2 (a^2 - f^2) \frac{1}{r_1^2} \right\} + \frac{\rho n^2}{\kappa + 1} \left\{ (a^2 - f^2)^2 \frac{1}{r_1^4} - (a^4 - f^4) \frac{1}{r_1^2} \right\},$$

$$\text{where} \quad r_1^2 = a^2 + f^2 + 2af \sin \theta, \quad (3.26)$$

whence, writing $f/a = d$, and using (1.4) and (3.25)

$$\bar{\theta}\bar{\theta}/\bar{\theta}\bar{\theta}_0 = 1 + 2 \frac{a^2}{r_1^2} d^2 (1 - d^2) - \left(\frac{1 + \eta}{1 - \eta} \right) \left\{ (1 - d^4) \frac{a^2}{r_1^2} - (1 - d^2)^2 \frac{a^4}{r_1^4} \right\}. \quad (3.27)$$

The curious fact emerges that for $d = 1$, $\theta \neq -\pi/2$ we have $\bar{\theta}\bar{\theta} = \bar{\theta}\bar{\theta}_0$, i.e. the peripheral stress when the axis of rotation passes through the boundary of the disk is constant and equal to the peripheral stress in the disk when it rotates about its centre, except at the axis itself where the stress becomes infinite. The variation of $\bar{\theta}\bar{\theta}/\bar{\theta}\bar{\theta}_0$ in general for the case $\eta = 0.25$ is shown in figure 2, the curves labelled 0, 1, 2 ... 10 corresponding respectively to $d = 0, 0.1, 0.2 \dots 1.0$. It will be seen that in the quadrants remote from the

axis the peripheral tension is fairly constant, but in the remaining quadrants it increases rapidly, after an initial decrease, the more closely we approach the axis. For values of d greater than about 0.82 this initial decrease is so large that for a small region of these quadrants the peripheral stress is compressive, as can be seen from curve 9 in figure 2, for which $d = 0.9$.

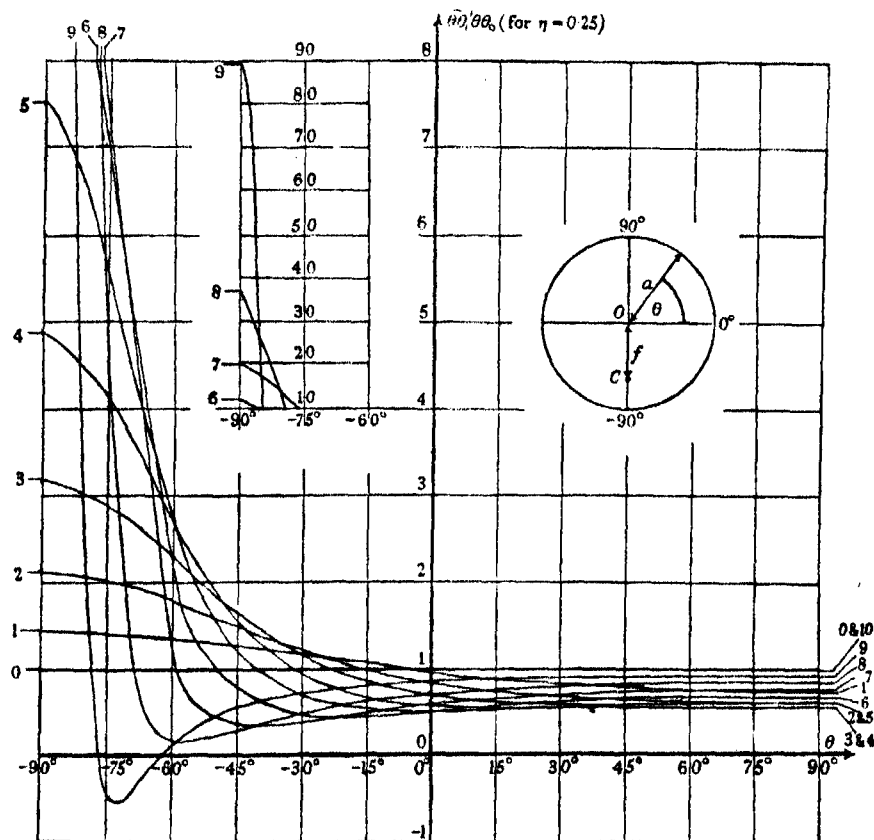


FIGURE 2. The peripheral stress (note the reduced scale for the inset portions of curves 6-9).

REFERENCES

- Jeffery 1921 *Phil. Trans. A*, **221**, 265.
 Mindlin 1938 *J. Appl. Phys.* p. 714.
 Stevenson 1945 *Proc. Roy. Soc. A*, **184**, 129.
 Timoshenko 1934 *Theory of Elasticity*. McGraw Hill Book Co.

Stress systems in isotropic and aeolotropic plates. V

By A. E. GREEN

(Communicated by G. R. Goldsbrough, F.R.S.—Received 19 October 1942

—Revised 8 June 1943)

A general theoretical solution is obtained for certain stress distributions in isotropic and aeolotropic plates containing holes of various types. The solution includes as special cases some well-known results for isotropic materials, and it is used here to obtain new results for both isotropic and aeolotropic plates. Numerical results are given for the distribution of stress round the edges of an elliptical hole in a spruce plank under tension, a square hole with rounded corners in an isotropic tension member and in an isotropic plate under shear, and a triangular hole with rounded corners in an isotropic tension member.

INTRODUCTION

1. The effect of a circular or elliptical hole on stress distributions in an isotropic plate has been extensively studied both theoretically and experimentally, and it has been found that the holes cause considerable local increases in the stresses. Experimental results have also been obtained for practical problems in which holes of other shapes affect the stress distributions (see Coker & Filon), but no corresponding theoretical results have been obtained so far, although a method of solution for such problems has been given by Muschelisvili (1933). In the first part of this paper a theoretical solution is given for problems of stress distributions in an infinite isotropic plate which contains a hole of a fairly general shape. Well-known results for an elliptical hole are easily deduced as special cases. The solution is then used to discuss certain stress distributions near rectangular holes and near holes represented by the parametric equations

$$x = na \cos \xi + b \cos n\xi, \quad y = -na \sin \xi + b \sin n\xi. \quad (1.1)$$

The family of curves (1.1) has been studied by D. M. Wrinch in connexion with electrostatic problems, and the circle and the ellipse are included in this family as special cases. The curve for which $n = 3$ and $a = 3b$ is of special interest as it represents approximately a square with rounded corners. Some numerical results are given for this case and also for the case $n = 2$, $a = 2b$ which represents approximately an equilateral triangle with rounded corners.

In the second part of this paper the general solution is extended so as to apply to aeolotropic plates. Numerical results are given for some stress distributions in certain specimens of spruce wood containing elliptical holes.

The method of solution used for both isotropic and aeolotropic plates is an extension of that given in a previous paper (Green 1942) which dealt with stresses near a circular hole. The solution depends on finding a conformal transformation of the form

$$z = F(\zeta), \quad (1.2)$$

where $z (= x + iy)$ is the plane of the plate and where $\zeta (= \xi + i\eta)$ takes real values at the edge of the hole. The point $\eta \rightarrow \infty$ corresponds to the point at infinity in the z -plane. The differential form of (1.2) will be needed and may be written

$$\frac{dz}{d\zeta} = F'(\zeta) = J e^{i\phi}, \quad (1.3)$$

where ϕ is the angle between the tangent to the curve $\eta = \text{constant}$ through any point and the x -axis. Transformations of this type have been used extensively by R. Morris and others for general harmonic problems.

When the hole is circular and of radius a the transformation is

$$z = ae^{-i\zeta}. \quad (1.4)$$

For an elliptical hole whose semi-axes are a' and b' the appropriate transformation is

$$z = c \cos(\zeta + i\alpha), \quad (1.5)$$

where c and α are real and positive and where $a' = c \cosh \alpha$, $b' = c \sinh \alpha$.

A rectangular hole is given by the transformation

$$\frac{dz}{d\zeta} = F'(\zeta) = (\cos 2\zeta - \cos 2\beta)^{\frac{1}{2}}. \quad (1.6)$$

When the hole is one of the family (1.1), the required transformation is

$$z = nae^{-i\zeta} + be^{in\zeta}, \quad (1.7)$$

where for the present b is restricted to be less than a . When b and a are equal the curve is star-shaped with $(n+1)$ cusps; otherwise it has continuous curvature. For some problems it is convenient to change the orientation of the hole of type (1.1) and to use the transformation

$$z = e^{i\pi n}(nae^{-i\zeta} + be^{in\zeta}). \quad (1.8)$$

Then, for the case $a = 3b$, $n = 3$, this represents a hole which is approximately a square with rounded corners, whose sides are parallel to the coordinate axes. The two forms (1.7) and (1.8) are, for convenience, included in the differential form

$$\frac{dz}{d\xi} = a_0 e^{-i\xi} + b_n e^{in\xi}. \quad (1.9)$$

SOLUTION FOR ISOTROPIC MATERIALS

2. The mean stresses in two dimensions may be expressed in terms of a stress-function χ by the equations

$$\bar{x}\bar{x} = \frac{\partial^2 \chi}{\partial y^2}, \quad \bar{y}\bar{y} = \frac{\partial^2 \chi}{\partial x^2}, \quad \bar{x}\bar{y} = -\frac{\partial^2 \chi}{\partial x \partial y}, \quad (2.1)$$

provided that $\nabla_1^4 \chi = 0. \quad (2.2)$

Attention will be confined to stress distributions which produce zero force resultants at the edge of the hole, the conditions for single-valued displacements being simple for these problems. For such distributions a general solution of (2.2) may be put in the form

$$\chi = f(z) + \bar{z}g(z), \quad (2.3)$$

where $\bar{z} = x - iy$ and where $f'(z)$ and $g(z)$ are regular functions of z which tend to zero at infinity (Green 1942). Only the real part of (2.3) is to be used. Hence, from (2.1), the stresses are found to be the real parts of

$$\left. \begin{aligned} \bar{x}\bar{x} &= -f''(z) + 2g'(z) - \bar{F}(\bar{\xi})g''(z), \\ \bar{y}\bar{y} &= f''(z) + 2g'(z) + \bar{F}(\bar{\xi})g''(z), \\ \bar{x}\bar{y} &= -if''(z) - i\bar{F}(\bar{\xi})g''(z), \end{aligned} \right\} \quad (2.4)$$

where dashes denote differentiation with respect to z , and where a bar placed over a function denotes the complex conjugate of the function.

The equations $\xi = \text{constant}$, $\eta = \text{constant}$ represent a net of orthogonal curves in the z -plane, and the stresses in the orthogonal coordinates (ξ, η) are related to the stresses in the cartesian coordinates (x, y) by the equations

$$\left. \begin{aligned} \xi\xi &= \bar{x}\bar{x} \cos^2 \phi + \bar{x}\bar{y} \sin 2\phi + \bar{y}\bar{y} \sin^2 \phi, \\ \eta\eta &= \bar{x}\bar{x} \sin^2 \phi - \bar{x}\bar{y} \sin 2\phi + \bar{y}\bar{y} \cos^2 \phi, \\ \xi\eta &= \frac{1}{2}(\bar{y}\bar{y} - \bar{x}\bar{x}) \sin 2\phi + \bar{x}\bar{y} \cos 2\phi. \end{aligned} \right\} \quad (2.5)$$

Thus, using (2.4),

$$\left. \begin{aligned} \xi\bar{\xi} &= 2g'(z) - F'(\zeta)\{f''(z) + \bar{F}(\bar{\zeta})g''(z)\}/\bar{F}'(\bar{\zeta}), \\ \eta\bar{\eta} &= 2g'(z) + F'(\zeta)\{f''(z) + \bar{F}(\bar{\zeta})g''(z)\}/\bar{F}'(\bar{\zeta}), \\ \xi\eta &= -iF'(\zeta)\{f''(z) + \bar{F}(\bar{\zeta})g''(z)\}/\bar{F}'(\bar{\zeta}), \end{aligned} \right\} \quad (2.6)$$

where the real parts of these expressions are to be taken.

Two functions of ζ , $V(\zeta)$ and $W(\zeta)$ are now introduced which are such that $V(\zeta)$ and $W(\zeta)$ are finite at infinity and such that the real part of $V(\zeta) = -\eta\eta_e$, and the imaginary part of $W(\zeta) = \xi\eta_e$ on the edge of the hole $\eta = 0$. $\eta\eta_e$ and $\xi\eta_e$ represent the values of the normal and shear stresses at the edge of the hole. From (2.6) it is seen that $V(\zeta)$ and $W(\zeta)$ may be written in the form

$$\left. \begin{aligned} V(\zeta) &= -2g'(z) - F'(\zeta)\{f''(z) + \bar{F}(\bar{\zeta})g''(z)\}/\bar{F}'(\bar{\zeta}), \\ W(\zeta) &= F'(\zeta)\{f''(z) + \bar{F}(\bar{\zeta})g''(z)\}/\bar{F}'(\bar{\zeta}). \end{aligned} \right\} \quad (2.7)$$

After some simplification it will be seen from (2.6) and (2.7) that the stresses may now be taken as the real parts of

$$\left. \begin{aligned} \xi\bar{\xi} &= -V(\zeta) - W(\zeta) - W(\zeta)\bar{F}'(\bar{\zeta})/\bar{F}'(\bar{\zeta}) - \frac{1}{2}\{V'(\zeta) + W'(\zeta)\}\{\bar{F}(\bar{\zeta}) - \bar{F}'(\bar{\zeta})\}/\bar{F}'(\bar{\zeta}), \\ \eta\bar{\eta} &= -V(\zeta) - W(\zeta) + W(\zeta)\bar{F}'(\bar{\zeta})/\bar{F}'(\bar{\zeta}) + \frac{1}{2}\{V'(\zeta) + W'(\zeta)\}\{\bar{F}(\bar{\zeta}) - \bar{F}'(\bar{\zeta})\}/\bar{F}'(\bar{\zeta}), \\ \xi\eta &= -iW(\zeta)\bar{F}'(\bar{\zeta})/\bar{F}'(\bar{\zeta}) - \frac{1}{2}i\{V'(\zeta) + W'(\zeta)\}\{\bar{F}(\bar{\zeta}) - \bar{F}'(\bar{\zeta})\}/\bar{F}'(\bar{\zeta}), \end{aligned} \right\} \quad (2.8)$$

where dashes attached to $V(\zeta)$ and $W(\zeta)$ denote differentiation with respect to ζ . The circumferential stress at the edge of the hole is the real part of the simple form

$$\xi\bar{\xi}_e = -V(\zeta) - 2W(\zeta). \quad (2.9)$$

Equations (2.8) represent a formal symbolical solution of the stress problem, the stresses being known when $V(\zeta)$, $W(\zeta)$, and the form of the transformation (1.2) are known. In the applications which follow attention is confined to those problems for which $\bar{F}'(\bar{\zeta})$ has zeros, but no poles, at points in the plate. Hence $V(\zeta)$ and $W(\zeta)$ will have poles at these zeros, but $V(\zeta) + W(\zeta) = -2g'(z)$ will be free from poles. Also, in general, $F'(\zeta)$ may be split up into two terms, the first tending to zero as η tends to infinity, and the second tending to infinity as η tends to infinity. The second term may usually be expressed in the form $a_0 e^{-i\zeta}$.

It will be noticed that when the hole is circular the solution readily reduces to that given previously (Green 1942) by using the transformation (1.4).

THE TENSION PROBLEM

3. Suppose that a uniform tension T is applied to the plate at infinity parallel to the x -axis. If the hole is absent the stresses in the coordinates (ξ, η) are found from the real parts of the expressions

$$\xi\xi = \frac{1}{2}T\{1 + F'(\zeta)/\bar{F}'(\bar{\zeta})\}, \quad \eta\eta = \frac{1}{2}T\{1 - F'(\zeta)/\bar{F}'(\bar{\zeta})\}, \quad \xi\eta = \frac{1}{2}iT F'(\zeta)/\bar{F}'(\bar{\zeta}). \quad (3.1)$$

Thus it is now necessary to find a stress system which tends to zero at infinity and which takes values at the edge of the hole which are the real parts of

$$-\eta\eta_e = \frac{1}{2}T\{1 - F'(\zeta)/\bar{F}'(\bar{\zeta})\}, \quad \xi\eta_e = -\frac{1}{2}iT F'(\zeta)/\bar{F}'(\bar{\zeta}). \quad (3.2)$$

Remembering the conditions which have been attached to the functions $F'(\zeta)$, $V(\zeta)$ and $W(\zeta)$, it will be seen that, for some transformations, $V(\zeta)$ and $W(\zeta)$ can be expressed in the form

$$\frac{2V(\zeta)}{T} = \frac{\bar{a}_0 e^{i\zeta} F'(\zeta) + \bar{F}'(\bar{\zeta}) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\bar{\zeta})} - \frac{\bar{a}_0 e^{i\zeta} \bar{F}'(\bar{\zeta}) + F'(\zeta) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\bar{\zeta})}, \quad (3.3)$$

$$\frac{2W(\zeta)}{T} = \frac{-\bar{a}_0 e^{i\zeta} F'(\zeta) + \bar{F}'(\bar{\zeta}) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\bar{\zeta})} + \frac{-\bar{a}_0 e^{i\zeta} \bar{F}'(\bar{\zeta}) + F'(\zeta) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\bar{\zeta})}. \quad (3.4)$$

The complete stress system can now be obtained from (2.8), (3.3) and (3.4), together with the stresses (3.1) which are transmitted from infinity. The stress $\xi\xi_e$ at the edge of the hole is of special interest and may readily be obtained from (2.9), (3.3) and (3.4) together with the stress transmitted from infinity. Thus

$$\xi\xi_e = -T + 2T(\bar{a}_0 e^{i\zeta} + a_0 e^{-i\zeta})/F'(\zeta), \quad (3.5)$$

where the real part only is to be taken.

When the transformation used is of the form (1.9) it is found that (3.3), (3.4) give a satisfactory solution provided n is equal to 0, 1 or 2, but not for $n \geq 3$, since for these cases the terms in $e^{2i\zeta}$ in $V(\zeta)$ and $W(\zeta)$ give infinite stresses at infinity. When $n \geq 3$ it is necessary to add extra terms to $V(\zeta)$ and $W(\zeta)$ in order to ensure that the stresses at infinity take the correct

values. The additional terms must be such that the ratio of the coefficients of $e^{2i\zeta}$ in the complete expressions for $V(\zeta)$, $W(\zeta)$ respectively is $(n-1)$, and the *extra* terms are

$$\frac{V(\zeta)}{T} = \frac{-(n-2)}{4} \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} - \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} \right) + \frac{(n-2)}{4} \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} + \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} \right) \left(\frac{\bar{a}_0 e^{i\zeta}}{\bar{F}'(\zeta)} - \frac{a_0 e^{-i\zeta}}{F'(\zeta)} \right) + \lambda \frac{\bar{a}_0 e^{i\zeta} F'(\zeta) + \bar{F}'(\zeta) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\zeta)}, \quad (3.6)$$

$$\frac{W(\zeta)}{T} = \frac{-(n-2)}{4} \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} + \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} \right) \left(\frac{\bar{a}_0 e^{i\zeta}}{\bar{F}'(\zeta)} + \frac{a_0 e^{-i\zeta}}{F'(\zeta)} - 1 \right) - \lambda \frac{\bar{a}_0 e^{i\zeta} F'(\zeta) - \bar{F}'(\zeta) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\zeta)}, \quad (3.7)$$

where

$$\lambda = \frac{(n-2)}{4} \left(1 - \frac{b_n \bar{b}_n}{na_0 \bar{a}_0} \right), \quad (3.8)$$

so that the *complete* expressions for $V(\zeta)$, $W(\zeta)$ are the sums of (3.3), (3.6) and (3.4), (3.7) respectively. Also the stress at infinity is now $(1+2\lambda)T$ parallel to the x -axis and the complete edge stress is the real part of

$$\xi \bar{\xi}_e = -T + 2T(\bar{a}_0 e^{i\zeta} + a_0 e^{-i\zeta})/F'(\zeta) + \frac{(n-2)T}{2} \left(\frac{\bar{a}_0 e^{i\zeta}}{\bar{F}'(\zeta)} + \frac{a_0 e^{-i\zeta}}{F'(\zeta)} - 1 \right) \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} + \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} + 1 - \frac{b_n \bar{b}_n}{na_0 \bar{a}_0} \right). \quad (3.9)$$

4. Using (1.5) it will be found that the stress at the edge of an elliptical hole, whose major axis is parallel to the tension at infinity, is given by the expression

$$\xi \bar{\xi}_e = T \frac{\sinh 2\alpha + 1 - e^{2\alpha} \cos 2\xi}{\cosh 2\alpha - \cos 2\xi}, \quad (4.1)$$

which agrees with results obtained by other writers.

When the hole is of the form (1.1), then the stress at its edge is obtained from (1.7) and (3.5) and is found to be

$$\xi \bar{\xi}_e = T \frac{a^2 - b^2 - 2a^2 \cos 2\xi + 2ab \cos (n-1)\xi}{a^2 + b^2 - 2ab \cos (n+1)\xi}, \quad (4.2)$$

provided $n \leq 2$. If $n \geq 3$ the edge stress is obtained from (1.7) and (3.9) and is

$$\begin{aligned} \frac{1}{2} \xi \bar{\xi}_e \{n - (n-2)b^2/(na^2)\} \{a^2 + b^2 - 2ab \cos (n+1)\xi\} / T \\ = \frac{1}{2}(a^2 - b^2) \{n - (n-2)b^2/(na^2)\} - \{na^2 - (n-2)b^2\} \\ \times \cos 2\xi + 2ab \cos (n-1)\xi, \end{aligned} \quad (4.3)$$

corresponding to a tension T at infinity parallel to the x -axis.

If the transformation (1.8) is used instead of (1.7), then the edge stress becomes

$$\xi \xi_e = T \frac{a^3 - b^2 - 2a^2 \sin 2\xi - 2ab \sin (n-1)\xi}{a^2 + b^2 - 2ab \cos (n+1)\xi}, \quad (4.4)$$

if $n \leq 2$, and

$$\begin{aligned} \frac{1}{2} \xi \xi_e \{n - (n-2)b^2/(na^2)\} \{a^2 + b^2 - 2ab \cos (n+1)\xi\} / T \\ = \frac{1}{2} (a^2 - b^2) \{n - (n-2)b^2/(na^2)\} - \{na^2 - (n-2)b^2\} \\ \times \sin 2\xi - 2ab \sin (n-1)\xi, \end{aligned} \quad (4.5)$$

if $n \geq 3$.

The stress at the edge of a rectangular hole which has two of its edges parallel to the tension at infinity is given by equations (1.6) and (3.5) and is

$$\left. \begin{aligned} \xi \xi_e &= -T + \frac{2\sqrt{2} T \cos \xi}{(\cos 2\xi - \cos 2\beta)^{\frac{1}{2}}} \quad (0 \leq \xi \leq \beta; 2\pi - \beta \leq \xi \leq 2\pi), \\ &= -T - \frac{2\sqrt{2} T \cos \xi}{(\cos 2\xi - \cos 2\beta)^{\frac{1}{2}}} \quad (\pi - \beta \leq \xi \leq \pi + \beta), \\ &= -T \quad (\beta < \xi < \pi - \beta; \pi + \beta < \xi < 2\pi - \beta). \end{aligned} \right\} \quad (4.6)$$

5. If now a uniform tension T is applied to the plate at infinity parallel to the y -axis, the stresses which are transmitted from infinity are found from the real parts of,

$$\xi \xi = \frac{1}{2} T \{1 - F'(\zeta)/\bar{F}'(\bar{\zeta})\}, \quad \eta \eta = \frac{1}{2} T \{1 + F'(\zeta)/\bar{F}'(\bar{\zeta})\}, \quad \xi \eta = -\frac{1}{2} i T F'(\zeta)/\bar{F}'(\bar{\zeta}). \quad (5.1)$$

The corresponding expressions for $V(\zeta)$ and $W(\zeta)$, for some transformations, are

$$\begin{aligned} \frac{2V(\zeta)}{T} &= \frac{\bar{a}_0 e^{i\zeta} F'(\zeta) + \bar{F}'(\bar{\zeta}) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\bar{\zeta})} \\ &\quad + \frac{\bar{a}_0 e^{i\zeta} \bar{F}'(\bar{\zeta}) + F'(\zeta) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\bar{\zeta})}, \end{aligned} \quad (5.2)$$

$$\begin{aligned} \frac{2W(\zeta)}{T} &= \frac{-\bar{a}_0 e^{i\zeta} F'(\zeta) + \bar{F}'(\bar{\zeta}) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\bar{\zeta})} \\ &\quad + \frac{\bar{a}_0 e^{i\zeta} \bar{F}'(\bar{\zeta}) - F'(\zeta) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\bar{\zeta})}, \end{aligned} \quad (5.3)$$

and, as before, the complete stress system may now be found. The stress at the edge of the hole takes the simple form

$$\xi \xi_e = -T + 2T(a_0 e^{-i\zeta} - \bar{a}_0 e^{i\zeta})/F'(\zeta), \quad (5.4)$$

where only the real part is to be taken.

When the transformation (1.9) is used (5.2) and (5.3) give the correct stress system if $n \leq 2$. When $n \geq 3$ the necessary *additional* terms are

$$\frac{V(\zeta)}{T} = \frac{(n-2)}{4} \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} - \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} \right) - \frac{(n-2)}{4} \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} + \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} \right) \left(\frac{\bar{a}_0 e^{i\zeta}}{\bar{F}'(\zeta)} - \frac{a_0 e^{-i\zeta}}{F'(\zeta)} \right) \\ + \lambda \frac{\bar{a}_0 e^{i\zeta} F'(\zeta) + \bar{F}'(\zeta) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\zeta)}, \quad (5.5)$$

$$\frac{W(\zeta)}{T} = \frac{(n-2)}{4} \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} + \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} \right) \left(\frac{\bar{a}_0 e^{i\zeta}}{\bar{F}'(\zeta)} + \frac{a_0 e^{-i\zeta}}{F'(\zeta)} - 1 \right) \\ - \lambda \frac{\bar{a}_0 e^{i\zeta} F'(\zeta) - \bar{F}'(\zeta) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\zeta)}, \quad (5.6)$$

and the stress at infinity is now $(1+2\lambda)T$ parallel to the y -axis. The complete edge stress for $n \geq 3$ is the real part of

$$\xi \xi_e = -T + 2T(a_0 e^{-i\zeta} - \bar{a}_0 e^{i\zeta})/F'(\zeta) \\ - \frac{(n-2)T}{2} \left(\frac{\bar{a}_0 e^{i\zeta}}{\bar{F}'(\zeta)} + \frac{a_0 e^{-i\zeta}}{F'(\zeta)} - 1 \right) \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} + \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} - 1 + \frac{b_n \bar{b}_n}{na_0 \bar{a}_0} \right). \quad (5.7)$$

6. The stress at the edge of an elliptical hole whose minor axis is parallel to the tension at infinity takes the well-known form

$$\xi \xi_e = T \frac{\sinh 2\alpha - 1 + e^{2\alpha} \cos 2\xi}{\cosh 2\alpha - \cos 2\xi}. \quad (6.1)$$

The stress at the edge of a hole given by the transformation (1.7) is

$$\xi \xi_e = T \frac{a^2 - b^2 + 2a^2 \cos 2\xi - 2ab \cos (n-1)\xi}{a^2 + b^2 - 2ab \cos (n+1)\xi}, \quad (6.2)$$

if $n \leq 2$, and

$$\frac{1}{2} \xi \xi_e \{n - (n-2)b^2/(na^2)\} \{a^2 + b^2 - 2ab \cos (n+1)\xi\}/T \\ = \frac{1}{2} (a^2 - b^2) \{n - (n-2)b^2/(na^2)\} + \{na^2 - (n-2)b^2\} \\ \times \cos 2\xi - 2ab \cos (n-1)\xi, \quad (6.3)$$

if $n \geq 3$, corresponding to a tension T at infinity parallel to the y -axis.

At the edge of a hole given by the transformation (1.8) the stress is

$$\xi \xi_e = T \frac{a^2 - b^2 + 2a^2 \sin 2\xi + 2ab \sin (n-1)\xi}{a^2 + b^2 - 2ab \cos (n+1)\xi}, \quad (6.4)$$

if $n \leq 2$, and

$$\frac{1}{2} \xi \xi_e \{n - (n-2)b^2/(na^2)\} \{a^2 + b^2 - 2ab \cos (n+1)\xi\}/T \\ = \frac{1}{2} (a^2 - b^2) \{n - (n-2)b^2/(na^2)\} + \{na^2 - (n-2)b^2\} \\ \times \sin 2\xi + 2ab \sin (n-1)\xi, \quad (6.5)$$

if $n \geq 3$.

For the rectangular hole (1.6) the edge stresses are now

$$\left. \begin{aligned} \xi \xi_s &= -T + \frac{2\sqrt{2} T \sin \xi}{(\cos 2\beta - \cos 2\xi)^{\frac{1}{2}}} & (\beta \leq \xi \leq \pi - \beta), \\ &= -T - \frac{2\sqrt{2} T \sin \xi}{(\cos 2\beta - \cos 2\xi)^{\frac{1}{2}}} & (\pi + \beta \leq \xi \leq 2\pi - \beta), \\ &= -T & (0 < \xi < \beta; \pi - \beta < \xi < \pi + \beta; 2\pi - \beta < \xi < 2\pi). \end{aligned} \right\} \quad (6.6)$$

THE SHEAR PROBLEM

7. In this section the plate is supposed to be subject to a uniform shear S , at large distances from the hole, parallel to the x - and y -axes. The stresses transmitted from infinity are then the real parts of

$$\xi \xi_s = -i S F'(\zeta) / \bar{F}'(\bar{\zeta}), \quad \eta \eta = i S F'(\zeta) / \bar{F}'(\bar{\zeta}), \quad \xi \eta = S F'(\zeta) / \bar{F}'(\bar{\zeta}), \quad (7.1)$$

and, for some transformations, the expressions for $V(\zeta)$ and $W(\zeta)$ are now

$$\frac{iV(\zeta)}{S} = \frac{\bar{a}_0 e^{i\zeta} \bar{F}'(\zeta) - F'(\zeta) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\zeta)}, \quad (7.2)$$

$$\frac{iW(\zeta)}{S} = \frac{\bar{a}_0 e^{i\zeta} \bar{F}'(\zeta) + F'(\zeta) \{F'(\zeta) - a_0 e^{-i\zeta}\}}{F'(\zeta) \bar{F}'(\zeta)}. \quad (7.3)$$

The stress at the edge of the hole reduces to the real part of

$$\xi \xi_s = 4i S \bar{a}_0 e^{i\zeta} / F'(\zeta). \quad (7.4)$$

The expressions (7.2) and (7.3) give the correct stress system for the transformation (1.9) if $n \leq 2$. When $n \geq 3$ the necessary *additional* terms are

$$\frac{iV(\zeta)}{S} = \frac{(n-2)}{2} \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} + \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} \right) - \frac{(n-2)}{2} \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} - \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} \right) \left(\frac{\bar{a}_0 e^{i\zeta}}{\bar{F}'(\zeta)} - \frac{a_0 e^{-i\zeta}}{F'(\zeta)} \right), \quad (7.5)$$

$$\frac{iW(\zeta)}{S} = \frac{(n-2)}{2} \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} - \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} \right) \left(\frac{\bar{a}_0 e^{i\zeta}}{\bar{F}'(\zeta)} + \frac{a_0 e^{-i\zeta}}{F'(\zeta)} - 1 \right), \quad (7.6)$$

and the shear stress at infinity is now $(1 + 2\lambda) S$. The complete edge stress for $n \geq 3$ is the real part of

$$\begin{aligned} \xi \xi_s &= 4i S \bar{a}_0 e^{i\zeta} / F'(\zeta) \\ &\quad + i(n-2) S \left(\frac{\bar{a}_0 e^{2i\zeta}}{a_0} - \frac{a_0 e^{-2i\zeta}}{\bar{a}_0} \right) \left(\frac{\bar{a}_0 e^{i\zeta}}{\bar{F}'(\zeta)} + \frac{a_0 e^{-i\zeta}}{F'(\zeta)} - 1 \right). \end{aligned} \quad (7.7)$$

8. The stress at the edge of an elliptical hole takes the known form

$$\xi \xi_c = \frac{2S e^{2\alpha} \sin 2\xi}{\cosh 2\alpha - \cos 2\xi}. \quad (8.1)$$

For a hole given by the form (1.7) the edge stress is

$$\xi \xi_c = \frac{4S\{a^2 \sin 2\xi + ab \sin (n-1)\xi\}}{a^2 + b^2 - 2ab \cos (n+1)\xi}, \quad (8.2)$$

if $n \leq 2$ and

$$\begin{aligned} \xi \xi_c \{n - (n-2)b^2/(na^2)\} \{a^2 + b^2 - 2ab \cos (n+1)\xi\} \\ = 4S\{\{na^2 - (n-2)b^2\} \sin 2\xi + 2ab \sin (n-1)\xi\}, \end{aligned} \quad (8.3)$$

when $n \geq 3$, corresponding to a uniform shear S at infinity.

For the hole (1.8) the stress is

$$\xi \xi_c = \frac{4S\{-a^2 \cos 2\xi + ab \cos (n-1)\xi\}}{a^2 + b^2 - 2ab \cos (n+1)\xi}, \quad (8.4)$$

if $n \leq 2$, and

$$\begin{aligned} \xi \xi_c \{n - (n-2)b^2/(na^2)\} \{a^2 + b^2 - 2ab \cos (n+1)\xi\} \\ = 4S\{-\{na^2 - (n-2)b^2\} \cos 2\xi + 2ab \cos (n-1)\xi\}, \end{aligned} \quad (8.5)$$

for $n \geq 3$.

The stress at the edge of the rectangular hole (1.6) is given by the following expressions:

$$\left. \begin{aligned} \xi \xi_c &= -\frac{2\sqrt{2} S \sin \xi}{(\cos 2\xi - \cos 2\beta)^{\frac{1}{2}}} & (0 \leq \xi \leq \beta; 2\pi - \beta \leq \xi \leq 2\pi), \\ &= \frac{2\sqrt{2} S \sin \xi}{(\cos 2\xi - \cos 2\beta)^{\frac{1}{2}}} & (\pi - \beta \leq \xi \leq \pi + \beta), \\ &= -\frac{2\sqrt{2} S \cos \xi}{(\cos 2\beta - \cos 2\xi)^{\frac{1}{2}}} & (\beta \leq \xi \leq \pi - \beta), \\ &= \frac{2\sqrt{2} S \cos \xi}{(\cos 2\beta - \cos 2\xi)^{\frac{1}{2}}} & (\pi + \beta \leq \xi \leq 2\pi - \beta). \end{aligned} \right\} \quad (8.6)$$

A numerical discussion of the above problems will be given at the end of the paper after the work has been extended so as to apply to aeolotropic materials.

AEOLOTROPIC MATERIALS

9. The material of the plate is supposed to have two directions of symmetry parallel to the coordinate axes (x, y). The mean stresses are still given by equations (2.1) and (2.5) but the stress function χ now satisfies the equation (Green & Taylor 1939)

$$\left(\frac{\partial^2}{\partial x^2} + \alpha_1 \frac{\partial^2}{\partial y^2} \right) \left(\frac{\partial^2}{\partial x^2} + \alpha_2 \frac{\partial^2}{\partial y^2} \right) \chi = 0, \quad (9.1)$$

where $\alpha_1 \alpha_2 = s_{11}/s_{22}, \quad \alpha_1 + \alpha_2 = (s_{66} + 2s_{12})/s_{22}, \quad (9.2)$

the elastic constants having their usual meanings. As in a previous paper (Green 1942) a general solution of equation (9.1) which gives single-valued expressions for the stresses and displacements may be taken to be the real part of

$$\chi = f(z + \gamma_1 \bar{z}) + g(z + \gamma_2 \bar{z}), \quad (9.3)$$

where $f'(z + \gamma_1 \bar{z})$ and $g'(z + \gamma_2 \bar{z})$ are regular functions of $z + \gamma_1 \bar{z}$ and $z + \gamma_2 \bar{z}$ respectively which tend to zero at infinity. Also

$$\gamma_1 = \frac{\alpha_1^{\frac{1}{2}} - 1}{\alpha_1^{\frac{1}{2}} + 1}, \quad \gamma_2 = \frac{\alpha_2^{\frac{1}{2}} - 1}{\alpha_2^{\frac{1}{2}} + 1}, \quad (9.4)$$

the analysis being confined to cases where γ_1 and γ_2 are real and numerically less than one.

Using (2.5) and (9.3), the stress system is found to be the real part of

$$\left. \begin{aligned} \widehat{x\bar{x}} &= - (1 - \gamma_1)^2 f''(z + \gamma_1 \bar{z}) - (1 - \gamma_2)^2 g''(z + \gamma_2 \bar{z}), \\ \widehat{y\bar{y}} &= (1 + \gamma_1)^2 f''(z + \gamma_1 \bar{z}) + (1 + \gamma_2)^2 g''(z + \gamma_2 \bar{z}), \\ \widehat{x\bar{y}} &= -i(1 - \gamma_1^2) f''(z + \gamma_1 \bar{z}) - i(1 - \gamma_2^2) g''(z + \gamma_2 \bar{z}), \end{aligned} \right\} \quad (9.5)$$

and

$$\left. \begin{aligned} \widehat{\xi\xi} &= - \left\{ \frac{\{F'(\xi) - \gamma_1 \bar{F}'(\bar{\xi})\}^2}{F'(\xi) \bar{F}'(\bar{\xi})} f''(z + \gamma_1 \bar{z}) - \frac{\{F'(\xi) - \gamma_2 \bar{F}'(\bar{\xi})\}^2}{F'(\xi) \bar{F}'(\bar{\xi})} g''(z + \gamma_2 \bar{z}) \right\}, \\ \widehat{\eta\bar{\eta}} &= \left\{ \frac{\{F'(\xi) + \gamma_1 \bar{F}'(\bar{\xi})\}^2}{F'(\xi) \bar{F}'(\bar{\xi})} f''(z + \gamma_1 \bar{z}) + \frac{\{F'(\xi) + \gamma_2 \bar{F}'(\bar{\xi})\}^2}{F'(\xi) \bar{F}'(\bar{\xi})} g''(z + \gamma_2 \bar{z}) \right\}, \\ \widehat{\xi\bar{\eta}} &= -i \left\{ \left(\frac{F'(\xi)}{\bar{F}'(\bar{\xi})} - \frac{\gamma_1 \bar{F}'(\bar{\xi})}{F'(\xi)} \right) f''(z + \gamma_1 \bar{z}) - \left(\frac{F'(\xi)}{\bar{F}'(\bar{\xi})} - \frac{\gamma_2 \bar{F}'(\bar{\xi})}{F'(\xi)} \right) g''(z + \gamma_2 \bar{z}) \right\}. \end{aligned} \right\} \quad (9.6)$$

Functions $V(\zeta)$ and $W(\zeta)$ may be defined as for an isotropic material and then, from (9.6), it will be seen that they may be put in the form

$$\left. \begin{aligned} V(\zeta) &= - \frac{\{F'(\zeta) + \gamma_1 \bar{F}'(\zeta)\}^2}{F'(\zeta) \bar{F}'(\zeta)} f''\{F(\zeta) + \gamma_1 \bar{F}(\zeta)\} \\ &\quad - \frac{\{F'(\zeta) + \gamma_2 \bar{F}'(\zeta)\}^2}{F'(\zeta) \bar{F}'(\zeta)} g''\{F(\zeta) + \gamma_2 \bar{F}(\zeta)\}, \\ W(\zeta) &= \left\{ \frac{F''(\zeta)}{\bar{F}'(\zeta)} - \frac{\gamma_1^2 \bar{F}''(\zeta)}{F'(\zeta)} \right\} f''\{F(\zeta) + \gamma_1 \bar{F}(\zeta)\} \\ &\quad + \left\{ \frac{F''(\zeta)}{\bar{F}'(\zeta)} - \frac{\gamma_2^2 \bar{F}''(\zeta)}{F'(\zeta)} \right\} g''\{F(\zeta) + \gamma_2 \bar{F}(\zeta)\}. \end{aligned} \right\} \quad (9.7)$$

General expressions may now be found for the value of the stress at any point of the plate, but these expressions will be very complicated and they are not likely to be of much practical value. The stress at the edge of the hole, which is of considerable interest, can, however, be expressed in a comparatively simple form. Thus, from (9.6) and (9.7), after some calculation, it is found that the edge stress is given by the real part of the formula

$$\begin{aligned} \xi \xi_e = & \frac{V(\zeta) \{(\gamma_1 + \gamma_2) F'(\zeta) \bar{F}'(\zeta) - F''^2(\zeta) - \gamma_1 \gamma_2 \bar{F}''^2(\zeta)\}}{\{F'(\zeta) + \gamma_1 \bar{F}'(\zeta)\} \{F'(\zeta) + \gamma_2 \bar{F}'(\zeta)\}} \\ & - \frac{2 W(\zeta) \{F''^2(\zeta) - \gamma_1 \gamma_2 \bar{F}''^2(\zeta)\}}{\{F'(\zeta) + \gamma_1 \bar{F}'(\zeta)\} \{F'(\zeta) + \gamma_2 \bar{F}'(\zeta)\}}. \end{aligned} \quad (9.8)$$

As before attention will be confined to those problems for which $\bar{F}'(\zeta)$ has zeros, but no poles, at points in the plate. Hence, from (9.7), it is seen that $V(\zeta)$ and $W(\zeta)$ will have poles at these zeros but that $V(\zeta) + W(\zeta)$ will be free from poles.

THE TENSION PROBLEM

10. When a uniform tension T is applied to the plate at infinity parallel to the x -axis the appropriate expressions for $V(\zeta)$ and $W(\zeta)$ are given in (3.3) and (3.4), and the corresponding stress at the edge of the hole is then found from (9.8). Thus, adding to this the stress transmitted from infinity, the total stress at the edge of the hole is found, after considerable reduction, to be given by the expression

$$\begin{aligned} \xi \xi_e \{F'(\zeta) + \gamma_1 \bar{F}'(\zeta)\} \{F'(\zeta) + \gamma_1 \bar{F}'(\zeta)\} \{F'(\zeta) + \gamma_2 \bar{F}'(\zeta)\} \{F'(\zeta) + \gamma_2 \bar{F}'(\zeta)\}^* \\ = (1 + \gamma_1)(1 + \gamma_2) T F'(\zeta) \bar{F}'(\zeta) [-F'(\zeta) \bar{F}'(\zeta) + (a_0 e^{-i\zeta} + \bar{a}_0 e^{i\zeta}) \{F'(\zeta) + \bar{F}'(\zeta)\} \\ + (\gamma_1 + \gamma_2) F'(\zeta) \bar{F}'(\zeta) + \gamma_1 \gamma_2 \{F''^2(\zeta) + \bar{F}''^2(\zeta) + F'(\zeta) \bar{F}'(\zeta) \\ - (a_0 e^{-i\zeta} + \bar{a}_0 e^{i\zeta}) \{F'(\zeta) + \bar{F}'(\zeta)\}\}]. \end{aligned} \quad (10.1)$$

When the transformation used is of the form (1.9) it can be verified that the expressions (3.3) and (3.4) for $V(\zeta)$ and $W(\zeta)$ give the correct stresses at infinity for $n = 0$ or 1 , so that (10.1) represents the correct edge stress for these cases. When $n \geq 2$ the infinity conditions are obscure so applications using these transformations are withheld at present.

When the hole is elliptical, with its major axis parallel to the tension at infinity, the edge stress is given by (1.5) and (10.1) and is found to be

$$\begin{aligned} & \xi_{\xi_0} \{ (1 + \gamma_1^2) \cosh 2\alpha + 2\gamma_1 - (1 + \gamma_1^2 + 2\gamma_1 \cosh 2\alpha) \cos 2\xi \} \\ & \quad \times \{ (1 + \gamma_2^2) \cosh 2\alpha + 2\gamma_2 - (1 + \gamma_2^2 + 2\gamma_2 \cosh 2\alpha) \cos 2\xi \} \\ & = T(1 + \gamma_1)(1 + \gamma_2) (\cosh 2\alpha - \cos 2\xi) \{ 1 + \gamma_1\gamma_2 + \frac{1}{2}(1 + \gamma_1 + \gamma_2 - \gamma_1\gamma_2) e^{2\alpha} \\ & \quad - \frac{1}{2}(1 - \gamma_1 - \gamma_2 - \gamma_1\gamma_2) e^{-2\alpha} - (\gamma_1 + \gamma_2 + e^{2\alpha} + \gamma_1\gamma_2 e^{-2\alpha}) \cos 2\xi \}. \quad (10.2) \end{aligned}$$

The value of the stress round a circular hole may be obtained from this formula by letting $\alpha \rightarrow \infty$, and the result agrees with that found in previous papers (Green & Taylor, 1945; Green 1942). Also (10.2) reduces to the formula (4.1) when $\gamma_1 = \gamma_2 = 0$, i.e. when the material is isotropic.

11. If a uniform tension T is applied to the plate at infinity parallel to the y -axis, then the values of $V(\zeta)$ and $W(\zeta)$ are given by (5.2) and (5.3), and the total stress at the edge of the hole can be found from the expression

$$\begin{aligned} & \xi_{\xi_0} \{ F'(\zeta) + \gamma_1 \bar{F}'(\zeta) \} \{ \bar{F}'(\zeta) + \gamma_1 F'(\zeta) \} \{ F'(\zeta) + \gamma_2 \bar{F}'(\zeta) \} \{ \bar{F}'(\zeta) + \gamma_2 F'(\zeta) \} \\ & = (1 - \gamma_1)(1 - \gamma_2) T F'(\zeta) \bar{F}'(\zeta) [- F'(\zeta) \bar{F}'(\zeta) + (\bar{a}_0 e^{i\zeta} - a_0 e^{-i\zeta}) \{ F'(\zeta) - \bar{F}'(\zeta) \} \\ & \quad - (\gamma_1 + \gamma_2) F'(\zeta) \bar{F}'(\zeta) + \gamma_1\gamma_2 \{ F'(\zeta) \bar{F}'(\zeta) - F'^2(\zeta) - \bar{F}'^2(\zeta) \\ & \quad - (\bar{a}_0 e^{i\zeta} - a_0 e^{-i\zeta}) \{ F'(\zeta) - \bar{F}'(\zeta) \} \}]. \quad (11.1) \end{aligned}$$

The stress at the edge of an elliptical hole whose minor axis is parallel to the tension at infinity can now be found from (1.5) and (11.1) and is

$$\begin{aligned} & \xi_{\xi_0} \{ (1 + \gamma_1^2) \cosh 2\alpha + 2\gamma_1 - (1 + \gamma_1^2 + 2\gamma_1 \cosh 2\alpha) \cos 2\xi \} \\ & \quad \times \{ (1 + \gamma_2^2) \cosh 2\alpha + 2\gamma_2 - (1 + \gamma_2^2 + 2\gamma_2 \cosh 2\alpha) \cos 2\xi \} \\ & = T(1 - \gamma_1)(1 - \gamma_2) (\cosh 2\alpha - \cos 2\xi) \{ -1 - \gamma_1\gamma_2 + \frac{1}{2}(1 - \gamma_1 - \gamma_2 - \gamma_1\gamma_2) e^{2\alpha} \\ & \quad - \frac{1}{2}(1 + \gamma_1 + \gamma_2 - \gamma_1\gamma_2) e^{-2\alpha} + (\gamma_1 + \gamma_2 + e^{2\alpha} + \gamma_1\gamma_2 e^{-2\alpha}) \cos 2\xi \}. \quad (11.2) \end{aligned}$$

THE SHEAR PROBLEM

12. If the plate is subject to a uniform shear S , at large distances from the hole, parallel to the x - and y -axes, the appropriate expressions for $V(\zeta)$

and $W(\zeta)$ are contained in (7.2) and (7.3). The total stress at the edge of the hole is then given by the real part of the formula

$$\begin{aligned} \xi \xi_0 \{ F'(\zeta) + \gamma_1 \bar{F}'(\zeta) \} \{ \bar{F}'(\zeta) + \gamma_1 F'(\zeta) \} \{ F'(\zeta) + \gamma_2 \bar{F}'(\zeta) \} \{ \bar{F}'(\zeta) + \gamma_2 F'(\zeta) \} \\ = 4iS(1 - \gamma_1 \gamma_2) F'(\zeta) \bar{F}'(\zeta) \{ \bar{a}_0 e^{i\zeta} \bar{F}'(\zeta) + (\gamma_1 + \gamma_2) \bar{a}_0 e^{i\zeta} F'(\zeta) \\ + \gamma_1 \gamma_2 [\bar{a}_0 e^{i\zeta} \bar{F}'(\zeta) + F'^2(\zeta)] \}. \quad (12.1) \end{aligned}$$

For an elliptical hole the edge stress becomes

$$\begin{aligned} \xi \xi_0 \{ (1 + \gamma_1^2) \cosh 2\alpha + 2\gamma_1 - (1 + \gamma_1^2 + 2\gamma_1 \cosh 2\alpha) \cos 2\xi \} \\ \times \{ (1 + \gamma_2^2) \cosh 2\alpha + 2\gamma_2 - (1 + \gamma_2^2 + 2\gamma_2 \cosh 2\alpha) \cos 2\xi \} \\ = 2S(1 - \gamma_1 \gamma_2) (\cosh 2\alpha - \cos 2\xi) (\gamma_1 + \gamma_2 + e^{2\alpha} + \gamma_1 \gamma_2 e^{-2\alpha}) \sin 2\xi. \quad (12.2) \end{aligned}$$

NUMERICAL DISCUSSION

13. Calculations are carried out for isotropic materials and for a specimen of spruce wood whose elastic constants are shown in table 1. These

TABLE 1

	s_{11}	s_{22}	s_{12}	s_{66}	α_1	α_2	γ_1	γ_2
spruce	15.5*	0.587	-0.33	11.5	16.91	1.56	0.608	0.111

values for spruce wood have been used in previous papers and they are chosen so that the grain of the wood is parallel to the y -axis. If the values of s_{11} and s_{22} are interchanged (which means a change in sign of γ_1 and γ_2), the grain is then parallel to the x -axis. The inverses of the constants s_{11} , s_{22} , s_{12} , s_{66} are measured in 10^3 kg./sq. mm.*

THE TENSION PROBLEM: ELLIPTICAL HOLE

14. The values of the stresses at the edge of an elliptical hole in a spruce plate under tension have been evaluated by using the formulae (10.2) and (11.2) together with (1.5) and (2.5), the semi-axes of the ellipse having the ratio $a':b' = 3:2$. Four cases are considered: the major axis of the ellipse may be either along or perpendicular to the tension and the tension may be either along or perpendicular to the grain. The results are given in tables 2 and 3. In both tables the coordinate $-\xi$ is the eccentric angle of the ellipse and $\xi = 0$ at one end of the major axis. The distribution of $\xi \xi_0$ over one quadrant of the hole is shown in figure 1 for the case where the tension is

* The units were incorrectly given as kg./sq.mm. in previous papers.

parallel to the grain and in figure 2 for the case when it is perpendicular to the grain. The values of $\xi\xi_c$ at the edge of a circular hole in a spruce plank are also shown in figures 1 and 2 for comparison. In these figures the sheet is supposed to be in a state of tension in the direction $\xi = 0$, so that when the tension is applied at right angles to the major axis of the hole the meaning attached to ξ is different from that used in tables 2 and 3.

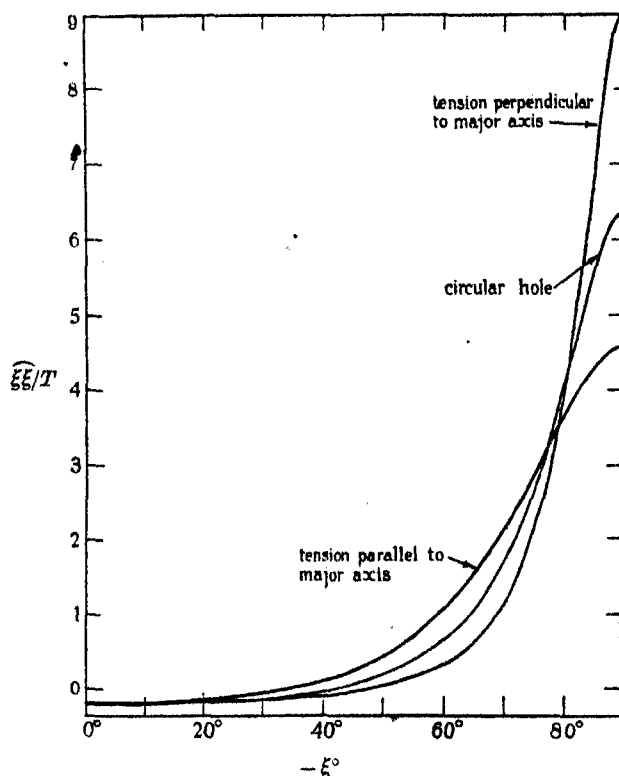


FIGURE 1. Stress distribution round the edges of elliptical holes in spruce; tension parallel to grain, parallel to $\xi = 0$.

From the figures and tables it is seen that the distribution of $\xi\xi_c$ round the elliptical holes in a spruce plank has the same general character as that round a circular hole. When the tension is applied parallel to the grain, the greatest variations in stress from that round a circular hole occur in the region of high stress concentration, but these regions are limited to small areas where the fibres which have been cut in making the hole lie close to the uncut fibres. When the tension is applied perpendicularly to the grain, the

greatest variations in stress from that round a circular hole occur mostly in the small regions near concentrations of high compressive stress.

In a previous paper (Green & Taylor 1945) an attempt was made to predict the type of failure which might be expected to occur in a spruce tension member containing a circular hole. The most interesting result was

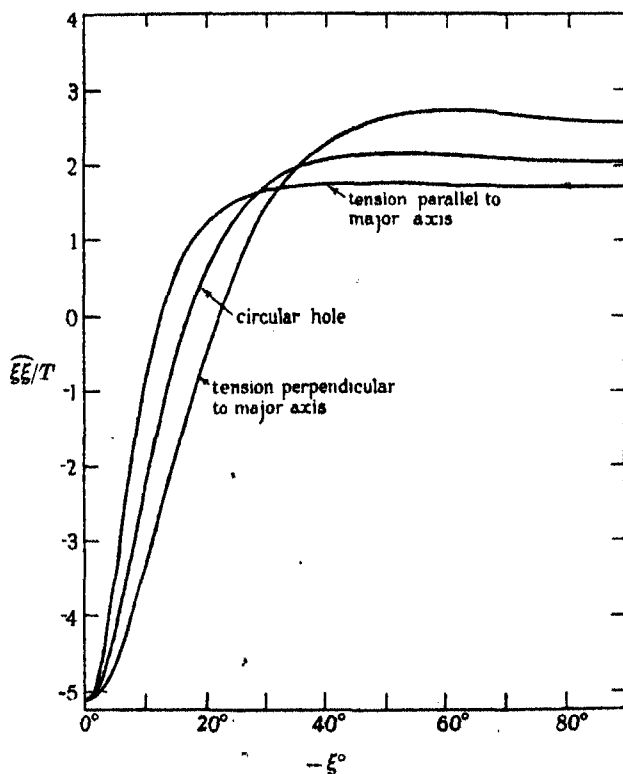


FIGURE 2. Stress distribution round the edges of elliptical holes in spruce; tension perpendicular to grain, parallel to $\xi = 0$.

that when the tension is applied parallel to the grain breakdown is to be expected by a shearing crack parallel to the grain at points on the circle which are about 12° from the position of maximum stress. A similar calculation for elliptical holes, using tables 2 and 3 and figures 1 and 2, for the case when the tension is applied parallel to the grain, and the major axis of the ellipse is either along or perpendicular to the grain, predicts the same kind of breakdown, but the angular position of the shear cracks is slightly different.

TABLE 2. VALUES OF STRESSES ON THE EDGE OF THE ELLIPSE:
TENSION PARALLEL TO THE MAJOR AXIS

$-\xi^\circ$	\widehat{xy}/T	\widehat{xx}/T	\widehat{yy}/T	$\widehat{\xi\xi}/T$
Tension parallel to grain				
0	0.0	0.0	-0.195	-0.195
10	0.0461	-0.0122	-0.174	-0.186
20	0.0645	-0.0352	-0.118	-0.153
30	0.0362	-0.0314	-0.0418	-0.0732
40	-0.0458	0.0576	0.0364	0.0940
50	-0.178	0.317	0.0993	0.417
60	-0.340	0.884	0.131	1.02
70	-0.476	1.96	0.116	2.08
75	-0.488	2.73	0.0871	2.82
80	-0.420	3.57	0.0494	3.62
85	-0.250	4.28	0.0146	4.30
90	0.0	4.57	0.0	4.57
Tension perpendicular to grain				
0	0.0	0.0	-5.13	-5.13
5	0.432	-0.0567	-3.29	-3.35
10	0.220	-0.0581	-0.830	-0.888
15	-0.179	0.0721	0.446	0.518
20	-0.505	0.276	0.925	1.20
30	-0.828	0.717	0.957	1.67
40	-0.858	1.08	0.681	1.76
50	-0.748	1.34	0.418	1.76
60	-0.581	1.51	0.224	1.73
70	-0.393	1.62	0.0953	1.71
80	-0.197	1.68	0.0232	1.70
90	0.0	1.70	0.0	1.70

THE TENSION PROBLEM: SQUARE HOLE

15. The stress at the edge of a rectangular hole in an isotropic tension member is given by the formulae (4.6) and (6.6). These formulae, however, only give practical values of the stress at points on the edge of the hole which are not near the corners of the rectangle, since at these corners the stress is theoretically infinite. It will be seen that the stress along most of the edge which is at right angles to the tension is compressive and equal to $-T$. For a square hole the stress at the middle points of the edges which are parallel to the tension is a tension and approximately equal to $1.82T$.

In order to get some idea of the effect of piercing a wall with rectangular openings Coker and Filon have, for simplicity, and to avoid stresses which would cause failure in the material, considered, by photo-elastic methods, the stress at the edge of a square hole whose angles are rounded off to a definite radius. It is probably possible to obtain an analytical description

TABLE 3. VALUES OF STRESSES ON THE EDGE OF THE ELLIPSE:
TENSION PERPENDICULAR TO THE MAJOR AXIS

$-\xi^\circ$	\widehat{xy}/T	\widehat{xx}/T	\widehat{yy}/T	$\widehat{\xi\xi}/T$
Tension parallel to grain				
0	0.0	0.0	9.04	9.04
5	-0.886	0.116	6.75	6.87
10	-0.949	0.251	3.59	3.84
15	-0.707	0.284	1.76	2.04
20	-0.464	0.253	0.850	1.10
30	-0.152	0.131	0.175	0.306
40	-0.0119	0.0149	0.00943	0.0244
50	0.0400	-0.0716	-0.0224	-0.0939
60	0.0502	-0.130	-0.0193	-0.150
70	0.0407	-0.168	-0.00988	-0.178
80	0.0221	-0.188	-0.00260	-0.191
90	0.0	-0.195	0.0	-0.195
Tension perpendicular to grain				
0	0.0	0.0	2.57	2.57
10	-0.644	0.170	2.43	2.60
20	-1.13	0.617	2.07	2.69
30	-1.36	1.17	1.57	2.74
40	-1.29	1.62	1.02	2.65
50	-0.958	1.71	0.536	2.25
60	-0.436	1.13	0.168	1.30
70	0.129	-0.530	-0.0312	-0.561
75	0.329	-1.84	-0.0588	-1.90
80	0.392	-3.33	-0.0460	-3.38
85	0.269	-4.61	-0.0157	-4.63
90	0.0	-5.13	0.0	-5.13

of the particular hole which they considered, but the subsequent calculation using the methods of this paper might prove tedious. The general character of the stress distribution round such a hole can, however, be indicated by using a hole represented by the simple expressions (1.1), taking the case $n = 3$; $a = 3b$, since this is approximately a square with the corners rounded off in a definite way.

For an isotropic material one case is considered numerically, namely that in which the tension is applied parallel to two sides of the square. The necessary results are either contained in formula (4.5) or they may be deduced from (4.3), (6.3) and (8.3). Putting $a = 3b$, $n = 3$, the edge stress becomes

$$\xi\xi_e = \frac{4T(1 - 2.7 \sin 2\xi)}{5 - 3 \cos 4\xi}, \quad (15.1)$$

and the numerical results are recorded in table 4. The point $\xi = 45^\circ$ corresponds to the mid-point of the side of the square which is perpendicular to the tension and the point $\xi = -45^\circ$ corresponds to the mid-point of the

side of the square which is parallel to the tension. The results are represented by a diagram in figure 3 in which tensional stress is denoted by lines drawn along the outward normals and compression is denoted by lines drawn along the inward normals.

Along the sides of the square which are parallel to the tension the stress is greatly above the value T and rises to a maximum of about $2.9T$ and then diminishes very rapidly and becomes an appreciable compression along the other boundary with a maximum at the centre of the line of $-0.85T$. These results are in general qualitative agreement with those found by Coker & Filon by photo-elastic experiments.

TABLE 4. VALUES OF THE STRESS AT THE EDGE
OF A SQUARE HOLE WITH ROUNDED CORNERS

ξ	45°	40°	35°	30°	25°	20°	15°
$\xi\xi$	-0.85	-0.849	-0.843	-0.824	-0.774	-0.657	-0.4
ξ	10°	5°	0°	-5°	-10°	-15°	-20°
$\xi\xi$	0.113	0.074	2	2.69	2.85	2.69	2.44
ξ	-25°	-30°	-35°	-40°	-45°		
$\xi\xi$	2.22	2.05	1.94	1.87	1.85		

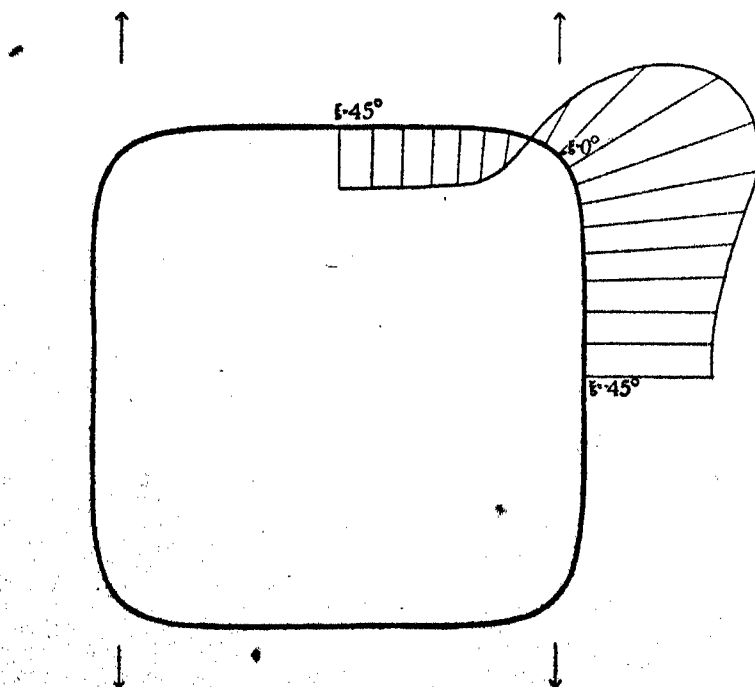


FIGURE 3. Distribution of stress round a square hole with rounded corners.

THE TENSION PROBLEM: TRIANGULAR HOLE

16. It is of some interest to find the variation of edge stress round a triangular hole in an isotropic tension member. As in the problem of the square hole the corners must be rounded off in order to avoid stresses which would cause breakdown of the material. The equation (1.1), when $a = 2b$, $n = 2$, represents approximately an equilateral triangular hole whose corners are rounded off, and from (4.2) it is seen that the stress at the edge of such a hole takes the simple form

$$\bar{\xi}_{\xi\xi} = \frac{T(3 - 8 \cos 2\xi + 4 \cos \xi)}{5 - 4 \cos 3\xi}, \quad (16.1)$$

when the tension is applied at right angles to a side, and from (6.2) the stress is given by

$$\bar{\xi}_{\xi\xi} = \frac{T(3 + 8 \cos 2\xi - 4 \cos \xi)}{5 - 4 \cos 3\xi}, \quad (16.2)$$

when the tension is applied parallel to a side.

TABLE 5. VALUES OF THE STRESS AT THE EDGE OF A TRIANGULAR HOLE WITH ROUNDED CORNERS

Tension perpendicular to a side							
ξ	0°	10°	20°	30°	40°	50°	60°
$\overline{\xi\xi}$	-1.0	-0.377	0.210	0.493	0.668	0.822	1
ξ	70°	80°	90°	100°	110°	120°	130°
$\overline{\xi\xi}$	1.24	1.60	2.2	3.27	5.05	5	1.18
ξ	140°	150°	160°	170°	180°		
$\overline{\xi\xi}$	-0.484	-0.893	-0.984	-0.999	-1.0		
Tension parallel to a side							
ξ	0°	10°	20°	30°	40°	50°	60°
$\overline{\xi\xi}$	7	4.28	1.79	0.707	0.189	-0.113	-0.333
ξ	70°	80°	90°	100°	110°	120°	130°
$\overline{\xi\xi}$	-0.531	-0.745	-1	-1.27	-1.15	1	2.72
ξ	140°	150°	160°	170°	180°		
$\overline{\xi\xi}$	2.48	2.09	1.84	1.71	1.67		

Numerical results are given in table 5 and the variation of edge stress is shown in figure 4 for the case when the tension is perpendicular to one side and in figure 5 for the case when the tension is parallel to a side. In the former case the stress at one corner is compressive and equal to $-T$. It

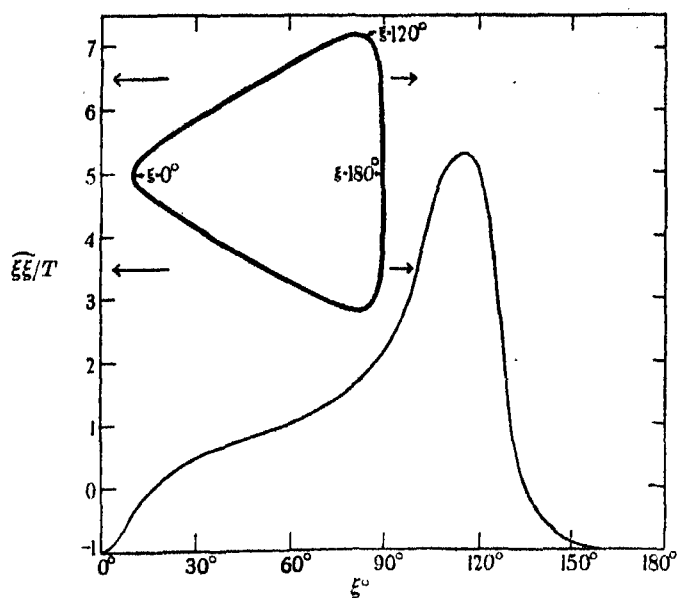


FIGURE 4. Stress distribution round the edge of a triangular hole. Tension perpendicular to a side.

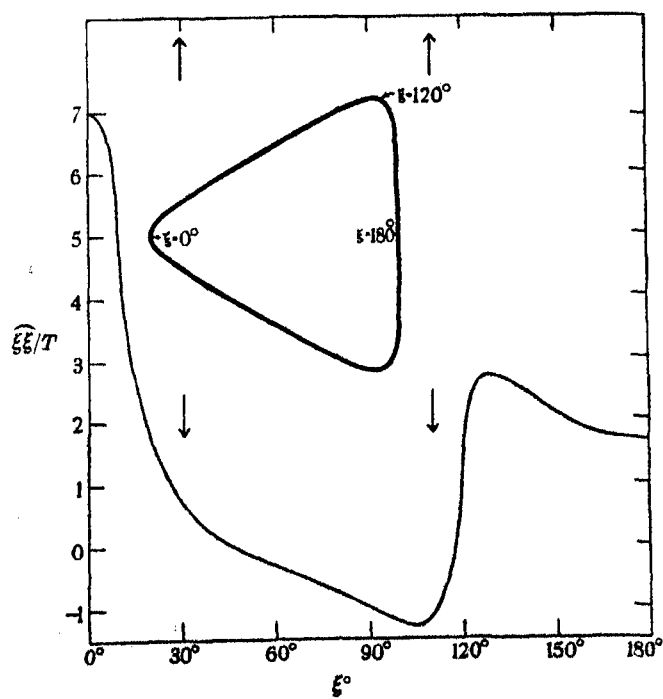


FIGURE 5. Stress distribution round the edge of a triangular hole. Tension parallel to a side.

then becomes a tension along the adjacent sides and at the other corners it rapidly increases to a maximum of $5.28T$ and then rapidly decreases and becomes a compression along most of the third side with a maximum of $-T$ at the centre of the side. In the latter case the stress is a maximum of $7T$ at one corner. It rapidly decreases along the adjacent sides and near the other corners becomes a compression with a maximum of $-1.3T$, and then rapidly increases to another maximum tension of $2.7T$ and finally falls to a tension of $1.67T$ at the mid-point of the third side.

THE SHEAR PROBLEM

17. Only one numerical example is considered here. A uniform shear S is applied to an isotropic plate which contains a square hole with rounded corners, the shear being applied parallel to the sides of the square. The formula for the edge stress is given by equation (8.5) when $a = 3b$ and $n = 3$, and it takes the simple form

$$\xi\xi_e = -\frac{13.5S \cos 2\xi}{5 - 3 \cos 4\xi}.$$

Some numerical values are given in table 6 where $\xi = 45^\circ$ is the mid-point of a side of the square and $\xi = 0$ is a corner of the square. It will be seen from the table that the edge stress reaches the high values $\pm 6.75S$ at the corners.

TABLE 6

ξ	45°	40°	35°	30°	25°	20°	15°	10°	5°	0
$\xi\xi_e$	0.0	0.300	0.633	1.04	1.67	2.31	3.34	4.70	6.10	6.75

REFERENCES

- Coker, E. G. & Filon, L. N. G. 1931 *A treatise on photo-elasticity*.
 Green, A. E. 1942 *Proc. Roy. Soc. A*, **180**, 173.
 Green, A. E. & Taylor, G. I. 1939 *Proc. Roy. Soc. A*, **173**, 162.
 Green, A. E. & Taylor, G. I. 1945 *Proc. Roy. Soc. A*, **184**, 181.
 Muschelisvili. 1933 *Z. angew. Math. Mech.* **13**, 264.

BAKERIAN LECTURE

On relaxation methods:* A mathematics for engineering science

BY R. V. SOUTHWELL, F.R.S.

(Delivered 17 June 1943—Received 3 August 1943)

1. By engineering, in this lecture, I intend the art whereby science is applied to useful ends; by engineering science, that corpus of knowledge—mathematics, physics, chemistry and the like—which is pursued with a view to such practical application. I am not concerned to defend these definitions, only to make my meaning clear. Such as they are, their distinction between engineering and engineering science is analogous with the distinction between clinical medicine—the art—and medicine—the science of our medical schools and research centres.

As engineering advances the scope of engineering science advances too, and roughly (I suppose) it may be said that the engineering science of any given time is the physics of fifty years before. Thus its field of study now is very much the same as that of nineteenth-century physicists like Kelvin, Stokes or Rayleigh: in 'field physics' (of which my lecture treats this afternoon) we are concerned with problems in hydrodynamics, elasticity and the like such as make up the bulk of their collected papers. There is, however, this difference in our outlook (and it arises because our science is directed to practical ends) that *we would rather have power to calculate approximately for any data than power to calculate exactly for data of a few restricted kinds.*

2. Take, as an example, Saint-Venant's well-known theory of torsion for a bar of non-circular cross-section. It is formally complete, and its equations have been satisfied exactly for various mathematical shapes—equilateral triangles, ellipses and the like. From a purely physical standpoint this is enough, and nineteenth-century physicists passed, in elasticity, on to other problems; but engineering science, in so far as it is concerned with the problem, is concerned with the torsional properties of shapes (e.g. of 'I-girders') which are not expressible mathematically, therefore

* The name 'Relaxation Methods' is an abbreviation of the more complete description given in the titles of my 1935 papers (Refs. 1 and 2): 'Stress-Calculation in Frameworks by the "Method of Systematic Relaxation of Constraints".'

are not tractable by orthodox mathematics. It has no great enthusiasm for exact solutions: 10 % accuracy in an estimation of stresses is good enough. But it does ask for methods which can be applied without restriction, *to any shape of cross-section*; and the mathematics which sufficed the nineteenth-century physicist it finds wanting in this respect.

Here and in other problems of field physics, for some six years I have been seeking, with a zealous team of co-workers* at no time numbering more than six, to furnish engineering science with a mathematics of its own; a mathematics not exact, but on the other hand not thus hedged about with troublesome restrictions. Being a team without official standing, it has fluctuated in size and personnel because ministries from time to time have taken its members for other work; but it has never been allowed to suffer extinction, since new recruits have been found.† From time to time our work has led to solutions having a war-time interest, and in consequence our energies have been focussed on a single objective, sometimes entailing much repetitive computation. In the intervals I have sought to extend the *range* of our methods, trying to guess what problems were most likely to attain war-time importance.

My aim to-day is not to explain the details of our methods, but to show the kind of thing that they can do. Figure 1, for example, shows one of our first solutions (in 1937) of the torsion problem. The equiangular section is, for orthodox analysis, one of the easiest to treat; but I do not think that orthodox analysis can do much with a *pierced* triangular section, and moreover, here and throughout this lecture it should be remembered that *any problem we have solved for one shape of boundary we could have solved for any*.

3. Having made that point, I now show further results of our earliest work. First, in the theory of torsion Prandtl (1923) sought to determine the consequences of 'yield' whereby, when the shear stress has attained some limiting value, the corresponding strain can increase without limit (figure 2). Here too a formal solution can be stated; but the difficulties of analysis are now much greater, because of the whole cross-section some parts behave 'elastically' and others 'plastically', and *the common boundary of the elastic*

* Their names appear as co-authors in the list of references. Acknowledgement is made to the Ministry of Supply, for grants in support of researches made on behalf of certain war committees; also to the Ministry of Aircraft Production (through its Aeronautical Research Committee) and to the Department of Scientific and Industrial Research, for grants in general support of our effort.

† In this connexion I gratefully acknowledge help received on many occasions from Dr C. P. Snow of the Central Register.

and plastic portions is not known in advance. Our computations for an equilateral triangle are shown in figures 3 *a*, *b*, where the contour curves show the direction and (by the closeness of their spacing) the intensity of the shear stress on cross-sections. On account of symmetry, only one-sixth of the complete triangle is reproduced in figure 3 *a*.

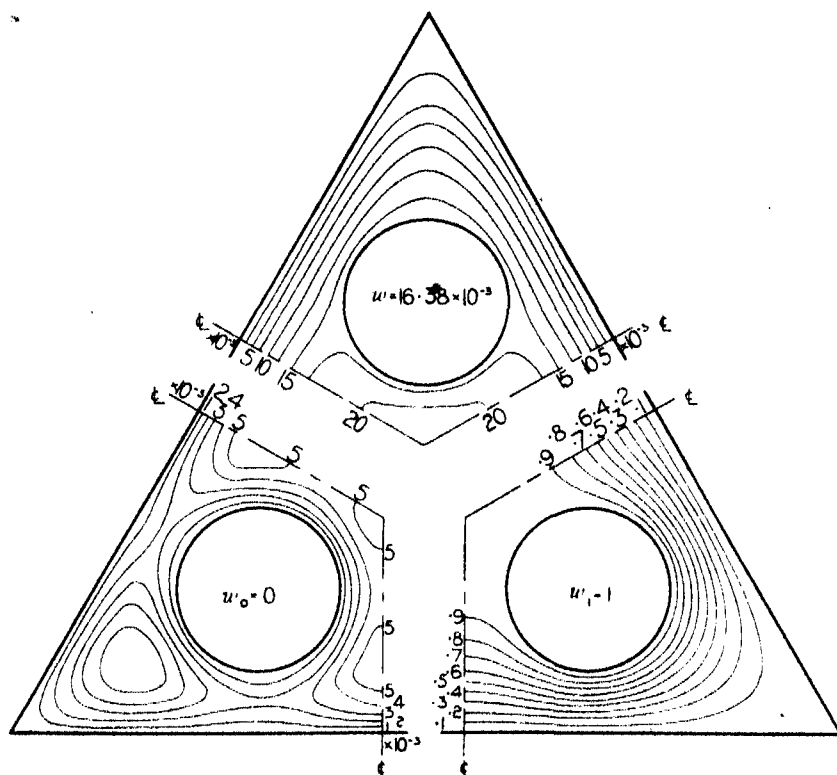


FIGURE 1. Solution of the torsion problem for a pierced triangle (Ref. 5, figure 7). The final solution (at top of diagram) was obtained by synthesis of the two solutions given below.

It will be seen that near the centre of each side the spacing of the contours is uniform so that the shear stress has a constant value. That value is the limiting stress f_Y in figure 2, and the dotted curve shows the extent of the plastic region (*not known initially*). As the twist increases, plastic strain extends over more and more of the cross-section; but however far it extends, elastic conditions are maintained in a spinal region extending from the centre to the corners (figure 3 *b*). Experiment gives general confirmation of these conclusions (cf. Nadai 1931, Chap. 19).

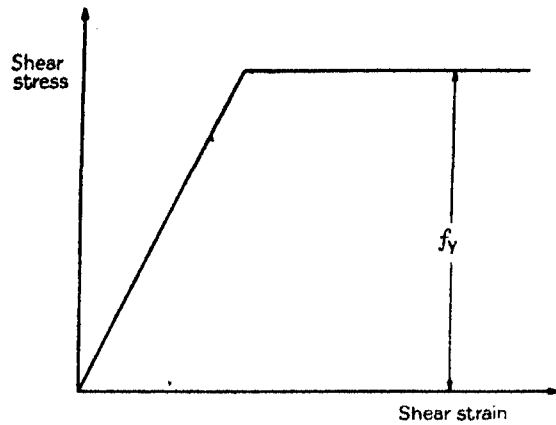
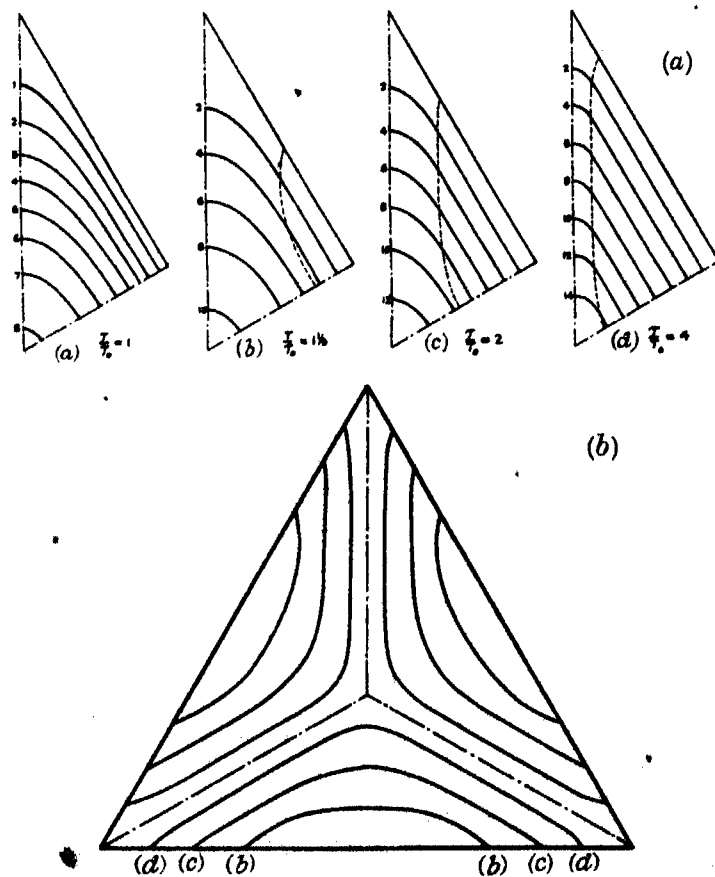


FIGURE 2. 'Plastic' stress-strain diagram (Ref. 5, figure 9).



FIGURES 3 a, b. 'Plastic torsion' of a triangular bar (Ref. 5, figures 10 and 11).

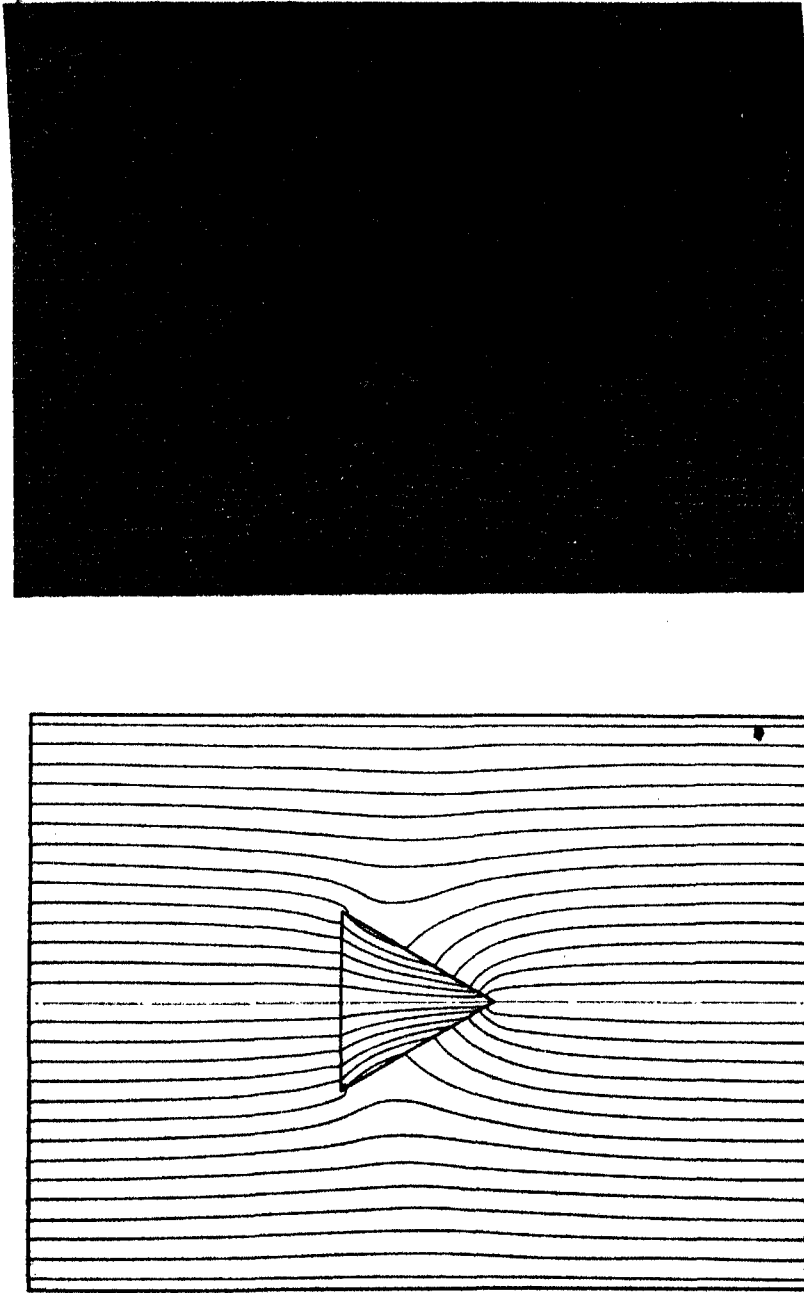


FIGURE 4. Magnetic lines (Ref. 5, figures 14 and 15), and comparison with figure 29 of Hele-Shaw & Hay (1900).

4. Secondly, we confront the phenomenon of refraction (e.g.) in the theory of magnetism when we deal with fields containing iron. Few cases of this kind yield to orthodox analysis, and Hele-Shaw & Hay (1900) devised a highly ingenious technique for obtaining solutions by experiment. Here again our methods have proved successful, and are not restricted to particular shapes of boundary. Figure 4 compares the results of computation with those of experiment, for an iron prism of triangular section.

5. A like freedom from restriction is found in dealing with the problem of conformal transformation: any of the four standard types of transformation (figure 5) can be effected by our methods for a region of any specified form. From the standpoint of engineering science, conformal transformation is a device whereby a problem hard to solve as presented can be simplified by a change of coordinates. Thus in studying the flow of a compressible fluid through a convergent-divergent nozzle we have found it advantageous to transform the 'field' of the fluid into a rectangle. Figure 6 shows our solution of this problem.

6. I have mentioned 'plastic torsion' as a problem hard to attack by orthodox methods for the reason that we cannot say *initially* what is the common boundary of the plastic and elastic regions: a similar difficulty is confronted in the treatment of fluid motion characterized by 'free surfaces'. Sometimes it can be turned by an *analytical* use of conformal transformation—as was shown by Kirchhoff and Rayleigh in their treatment of jets, etc., in two dimensions; but this elegant device has pitfalls—as was found by Davison & Rosenhead (1940) in a study of percolation through granular material (figure 7). Here, a single and simple boundary condition is imposed along the sides of the rectangular retaining wall, but the rest of the boundary (namely, the free surface of the 'water table' AE) is *not* known in advance, and on it a double boundary condition must be satisfied. Using the device, Davison & Rosenhead compelled this remaining boundary to start at A and to finish at D ; but having no means of allowing for the vertical boundary DE they could not compel it to stay within the porous material, and in fact it was found (in the absence of an assumed evaporation) to pass from left to right through the vertical boundary DE before returning (as on their assumptions it must) to D .

Later, when I come to describe our methods, you will see that they are essentially *tentative* like the engineering process of 'scraping' to a surface-plate or gauge: for that reason I felt confident that they would serve in cases where some part of a boundary is initially unknown, and applied to

this problem (in 1940-1) they led quickly to positive results. The true solution, as you see, involves 'seepage' of water through a part of the vertical side of the retaining wall.

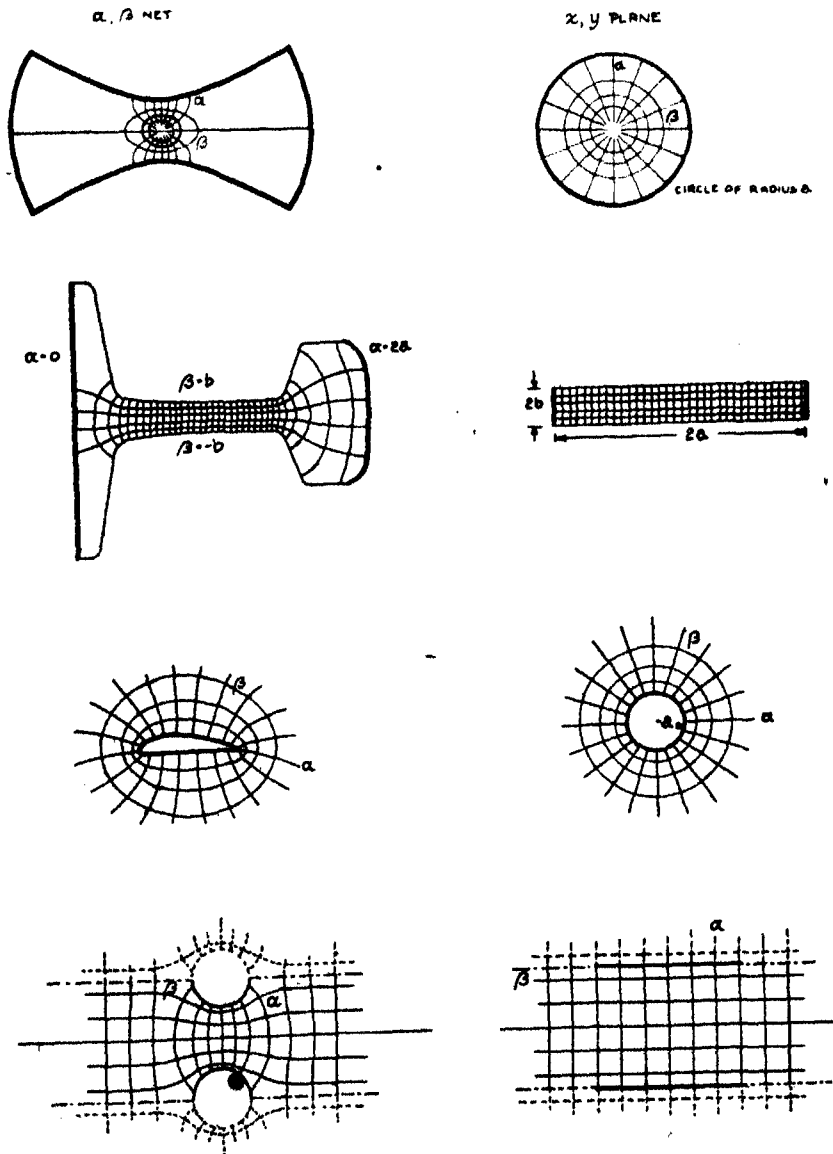
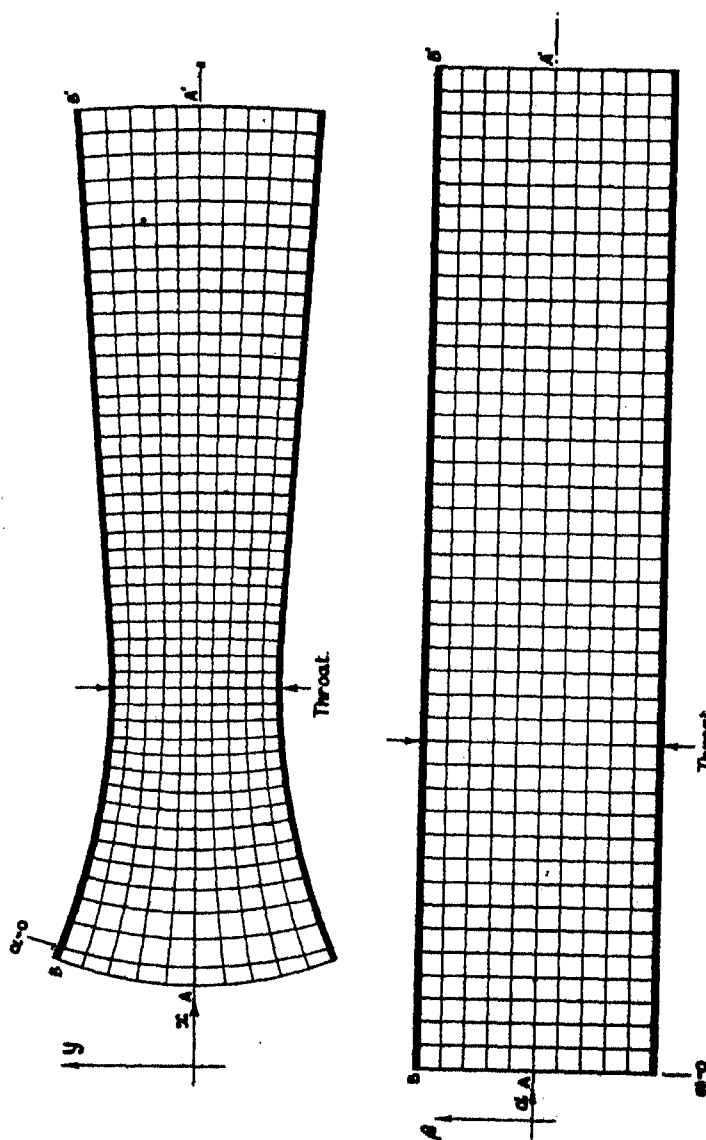


FIGURE 5. Types of conformal transformation (Ref. 7, figure 1).



Part of figure 6; legend on facing page.

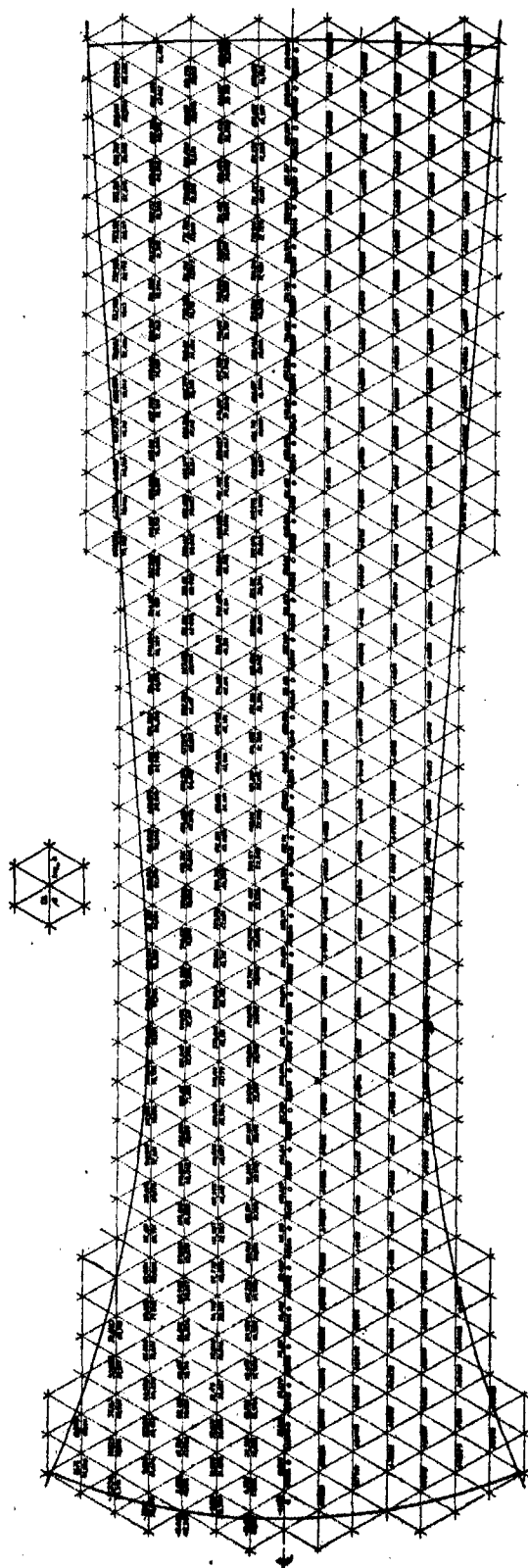


FIGURE 6. Conformal transformation of convergent-divergent nozzle (Ref. 16, figures 2 and 7a).

Figures 8 and 9 show further examples, and throw some light on the failure of conformal transformation in the case examined by Davison & Rosenhead. In both a rubble 'blanket' is assumed to be provided, as is

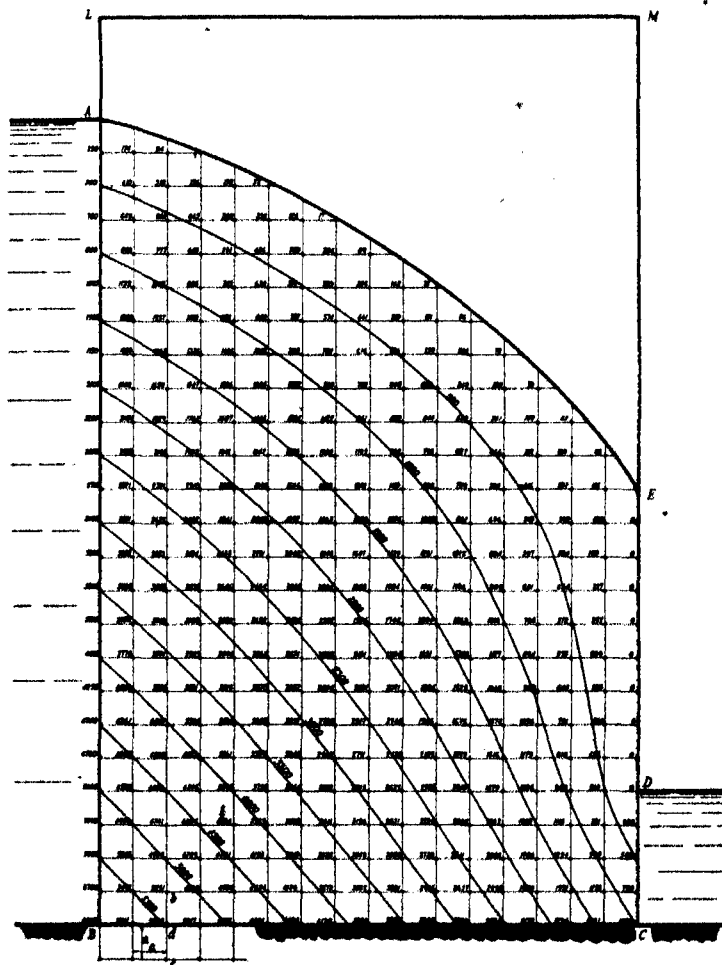


FIGURE 7. Percolation problem of Davison & Rosenhead (Ref. 9, figure 5).
The curves are contours of constant pressure.

customary,* to prevent erosion by drainage down the side exposed to air. In the second (due to the assumption of two strata of different porosities) refraction enters again as a complicating factor.

* Cf. Casagrande (1937).

7. So far I have dealt with problems differing in respect of their boundary conditions but all having as governing equation either the two-dimensional form of Laplace's equation, viz.

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right] \psi = 0, \quad (1)$$

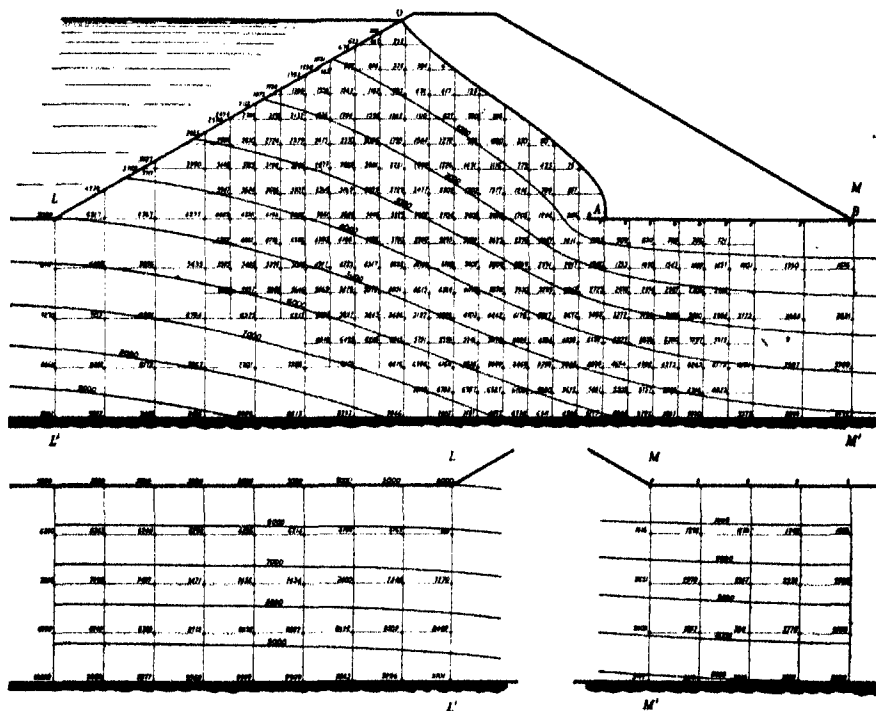


FIGURE 8. Percolation through a levee with rubble 'blanket' (Ref. 9, figure 6).
The curves are contours of constant pressure.

or the two-dimensional form of Poisson's equation, viz.

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right] w + Z(x, y) = 0, \quad (2)$$

where $Z(x, y)$ is specified. I now give some account of our methods as applied to these 'plane-harmonic' equations, treating (1) as a particular case of (2).

As my slides have shown, we present our solutions in the form of numerical values of the wanted function (w) at nodal points of a uniform lattice

or 'net'. The meshes (in theory) may be either hexagonal, or square, or triangular (figure 10); the values satisfy, not the governing equation (2) as it stands, but the approximation to it which results when its differentials are replaced by finite-difference approximations. Such replacement, of course, is no new device: indeed, it is hard to suggest an alternative, if the aim is to evolve a method applicable to any shape of boundary. I believe

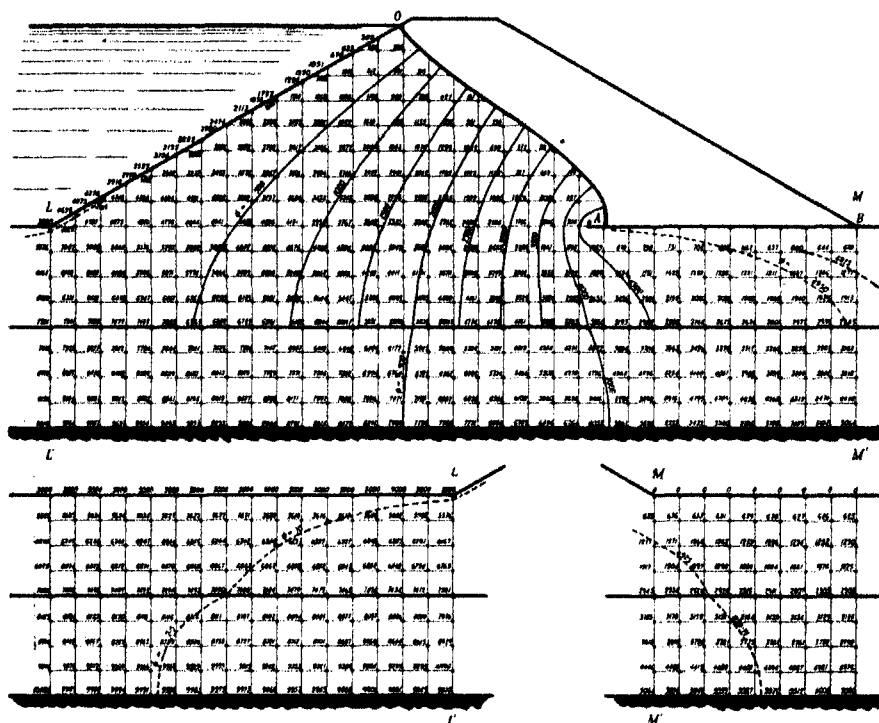


FIGURE 9. The same problem modified by the assumption of two strata having different porosities (Ref. 9, figure 7).

The curves are contours of 'velocity potential' ϕ .

that triangular nets (which have some advantages in respect of accuracy) have not been employed before; but they *could* have been employed by earlier investigators whose methods can hardly be described as 'relaxational', and it is not this feature that I want to stress.

8. What I regard as the essence of 'relaxation methods' is their visualization of any plane-harmonic problem as concerned with a mechanical system executing controlled displacements under the action of constraints. That

notion came to me first as a means of stress-calculation in engineering (structural) frameworks, where 'redundancy' (i.e. a superfluity of members above the number which would suffice to render a framework 'just stiff') introduces difficulties and uncertainties with which every engineer is familiar. Postponing for the moment my account of our attack on these structural problems, I now give a mechanical interpretation of our square and triangular nets.

Prandtl (1903) showed that an equation of the form of (2) governs (if these are small) the transverse deflexions w of a uniformly tensioned membrane (e.g. a soap film) under the action of transverse pressure having, at every point, an intensity proportional to Z . Griffith & Taylor (1917) utilized this analogy to find experimentally, with the aid of soap films, solutions of the torsion problem for cross-sections of non-mathematical

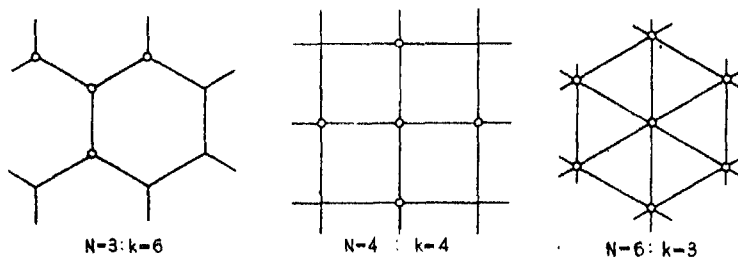


FIGURE 10. Three types of relaxation net (Ref. 5, figure 1).

shapes. We have shown (1943) that the finite-difference approximation to (2) may be interpreted similarly, as governing the transverse deflexions of nodal points of a tensioned net when the transverse pressure is concentrated at these points in accordance with the rules of statics (Ref. 14, §§ 7-9). Consequently an actual net might be used (in the manner of Griffith & Taylor) to solve the finite-difference equation by experiment; but in fact the net is tractable theoretically, and greater accuracy is attainable in calculation. (In experiment, you see, there is this dilemma—that unless considerable displacements are permitted there is nothing much to measure; whereas if they are, then the analogue is inexact,—the deflexions are governed by a different equation.)

9. A great advantage of the 'net analogy' (to minds like mine) is that it enables one to judge *by intuition* the extent of the error which is *and must be* entailed by our use of finite-difference approximations; and for that reason I prefer to spend my own energies in extending the range of problems which can be treated by this simple concept, notwithstanding that

the more mathematically minded of my colleagues are bent (I am glad to say) on improving the accuracy of our approximations.

The appeal to me of the mechanical ('net') analogy may be illustrated by reference to a slide shown earlier (figure 7). Here, if we state our problem as that of determining the fluid pressure p (which is 'plane-harmonic'),

- (i) the wanted function has specified values along AB , CDE , EA , and specified normal gradients along BC ,
- (ii) a further condition is imposed along AE ,

but the form of AE is not known initially, being determined by a double boundary condition. We made no progress until we examined this double condition as relating to p interpreted mechanically, as the transverse displacement of a membrane fixed (to sloping boundaries) along AB , BC , CD , unloaded except along AE , and there loaded by edge forces having uniform horizontal line-intensity; along AED the displacement must be zero. A model (figure 11) shows the nature of the mechanical problem. Every string is maintained at the same tension by a hanging load, and other weights exert an equal pressure where they touch the string. I move each weight until it can no longer keep the string pressed against the base (or 'datum plane'): when every weight is in equilibrium, I have the wanted curve.

We set ourselves to solve this problem by computation (of course, on a net finer than the net in my model).^{*} Then solutions came rapidly, *owing to the tentative nature of the relaxation process.*

10. I must now explain that process, and to do so I will summarize ideas by which (in 1935) I was led to a relaxational treatment of *braced frameworks*. I shall be very brief, for I have already given them in a book published late in 1940 (Ref. 19).

The standard problem in frameworks is, given the loads which come upon the joints, to deduce the resulting stresses or (what amounts to the same thing) the joint-displacements. Let us suppose that *displacements* are wanted. Then the orthodox procedure is, taking the joint-displacements as unknowns, to formulate equations of equilibrium for every joint, having the specified forces on their right-hand sides. Up to this point a firm grasp is kept on physical realities; but now there follows a series of operations, performed in accordance with definite mathematical rules, but in which all

^{*} The net was given large meshes in the model, partly for simplicity of manufacture and adjustment, but also in order to illustrate the satisfactory representation of a wanted function which even a coarse net permits. When the weights are in adjustment, the eye gains a quite clear impression (no doubt, helped largely by imagination) of a continuous function coinciding with the net at its nodal points.

physical contact is lost. We dive, so to speak, into a surf of computations, from which we emerge, slightly breathless, with a result which should be,

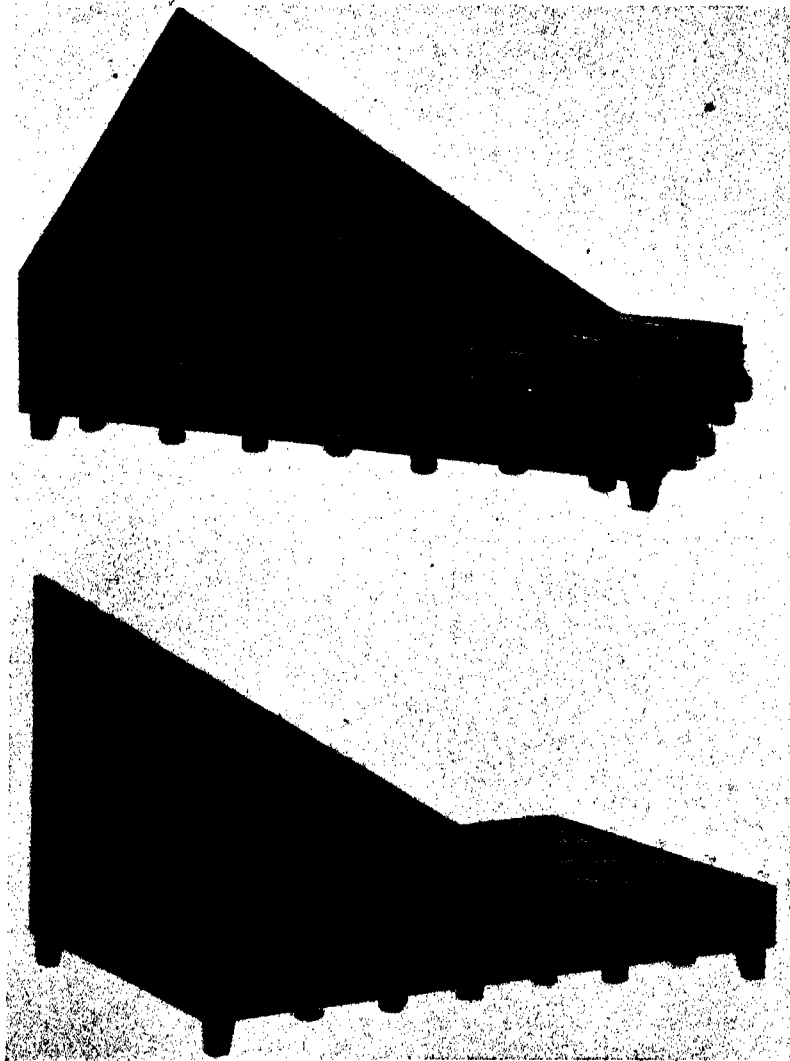


FIGURE 11. Model illustrating net analogue of figure 7.
(Photographs by Prof. C. M. White.)

and we hope is, a pearl. It is a series of displacement values, which ought to satisfy all of the original equations. We try them, and perhaps they do. If not, we must dive again.

In contrast with this process, I sought to reproduce in computation, step by step, a process which might (at least in imagination) be applied to the actual framework. "An ordinary engineering 'jack' (e.g. for automobiles) is a means whereby a controlled displacement may be imposed at any point: it is easy to imagine devices whereby this displacement can be recorded, together with the load sustained at any instant; also to visualize an arrangement in which every joint of a framework which would normally be free to move is provided with a jack of this kind to control its displacement. Suppose that initially the jacks fix the joints in positions such that the framework is not strained: then, plainly, when the external loads are applied they will be taken wholly by the jacks. Suppose that subsequently one jack is relaxed, so that one joint is permitted to travel slowly through a specified distance: then load will be transferred from that jack to adjacent jacks and to the framework, and strain-energy will be stored in the latter. If the force on the jack which is relaxed had a component in the direction of the travel, that jack will be relieved, and strain-energy will be stored in the framework at the expense of the potential energy of the external forces" (Ref. 19, § 3).

Here then is an imagined process whereby a framework could be brought from the unstrained to the fully-strained configuration; and every step is easy to follow in calculation. The governing consideration, you see, is that *indirect solutions are easy*: there is no difficulty in computing the forces brought into play when all but one particular joint are held fixed and on that joint a specified displacement is imposed. So, for every step in the loading process, we can compute the relative changes in the jack load-readings; then, in a further computation, we can follow the whole process, keeping touch with reality throughout. Two computational tables are entailed (figure 12): the first (the 'operations table') exhibits the effects on the recorded loads of displacing every joint severally, also (e.g. operations '6' and '7') of any simultaneous displacement of two or more joints which we may expect to find useful; in the second (the 'liquidation table') the first line shows the applied forces which are the loads initially recorded, subsequent lines show a steady decrease in magnitude of the recorded loads in consequence of operations which are detailed in the first and second columns. The tentative nature of the relaxation process will be apparent to anyone who examines this table carefully.

11. Now, in place of a braced framework, let us suppose that we load our tensioned net. Nothing need be altered in our imagined (physical) process, and again every step can be followed in computation. In fact

there is this simplification, that whereas in the framework every unit operation may be different, in the net (owing to its regularity) every operation is the same. We no longer need an operations table: it

TABLE 1
(Units: 1 ton weight; 1 foot.)

Operation no.	Nature of operation	(1) X_A	(2) X_B	(3) Y_A	(4) Y_B	(5) Y_C	X_C	X_D	Y_D
1 (a)	$u_A = 1$	-3,980	0	312	0	768	1,280	2,700	-1,080
1 (b)	$= 0.2513$	-1,000	0	78	0	193	323	678	-271
2 (a)	$u_B = 1$	0	-4,100	0	720	0	3,380	780	-720
2 (b)	$= 0.2439$	0	-1,000	0	175	0	825	175	-175
3 (a)	$v_A = 1$	312	0	-5,393	4,500	461	768	-1,080	432
3 (b)	$= 0.1854$	88	0	-1,000	834	85	142	-200	80
4 (a)	$v_B = 1$	0	720	4,500	-5,720	0	0	-720	720
4 (b)	$= 0.1916$	0	138	962	-1,000	0	0	-138	138
5 (a)	$v_C = 1$	768	0	460.8	0	-2,480.8	-768	0	2,020
5 (b)	$= 0.4031$	309	0	185	0	-1,000	-309	0	814
6 (a)	$v_A = v_B = 1$	312	720	-893	-720	461	768	-1,809	1,152
6 (b)	$= 0.6200$	193	446	-554	-446	286	476	-1,116	714
7 (a)	$v_A = -v_B = 1$	312	-720	-9,893	9,720	461	768	-380	-288
7 (b)	$= 0.0510$	16	-37	-504	496	23	39	-18	-15

Operation	Multiplier $\times 10^4$	X_A	X_B	Y_A	Y_B	Y_C
	(Initial forces)→	15	7.07	..	7.07	..
1 (b)	15	-15	0	1.18	0	2.89
		0	7.07	1.18	7.07	2.89
2 (b)	7	0	-7.00	0	-1.22	0
		0	0.07	1.18	8.29	2.89
6 (b)	10	1.93	4.46	-5.54	-4.46	2.86
		1.93	4.53	-4.36	3.83	5.75
7 (b)	-8	-0.13	0.30	4.03	-3.97	-0.18
		1.80	4.83	-0.33	-0.14	5.57
5 (b)	7	2.18	0	1.80	0	-7.00
		3.98	4.83	0.97	-0.14	-1.43
2 (b)	6	0	-6.00	0	1.08	0
		3.98	-1.17	0.97	+0.91	-1.43
1 (b)	8	-5.00	0	0.39	0	+0.97
		-1.02	-1.17	1.36	+0.91	-0.46
6 (b)	2.5	0.48	+1.12	-1.39	-1.11	0.72
		-0.64	-0.05	-0.03	-0.20	0.26
5 (b)	0.20	0.06	0	0.04	0	-0.20
		-0.48	-0.05	+0.01	-0.20	0.06
1 (b)	-0.5	0.60	0	-0.04	0	-0.10
		0.02	-0.05	-0.03	-0.20	-0.04
6 (b)	-0.25	-0.05	-0.11	+0.14	0.11	-0.07
		-0.03	-0.16	+0.11	-0.09	-0.11
2 (b)	-0.2	0	0.30	0	-0.03	0
		-0.03	0.04	0.11	-0.12	-0.11
7 (b)	0.2	0	-0.01	-0.10	0.10	0
		-0.03	0.03	0.01	-0.02	-0.11
5 (b)	-0.12	-0.04	0	-0.02	0	0.12
		-0.07	0.08	-0.01	-0.02	+0.01
1 (b)	-0.06	0.06	0	0	0	-0.02
		-0.01	0.03	-0.01	-0.02	-0.01
6 (b)	-0.04	0	-0.02	0.02	0.020	-0.01
		-0.01	0.01	0.01	0	-0.02

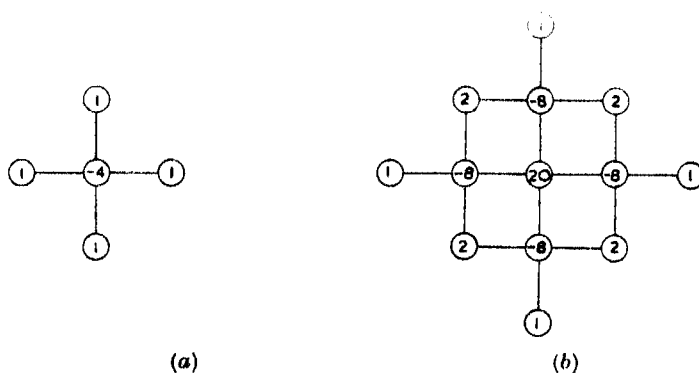
FIGURE 12. Typical 'operations' and 'liquidation' table
(Ref. 19, tables II and III).

is replaced by a standard 'relaxation pattern' (figure 13). And liquidation, too, no longer calls for a table: it can be effected on the 'relaxation net'.

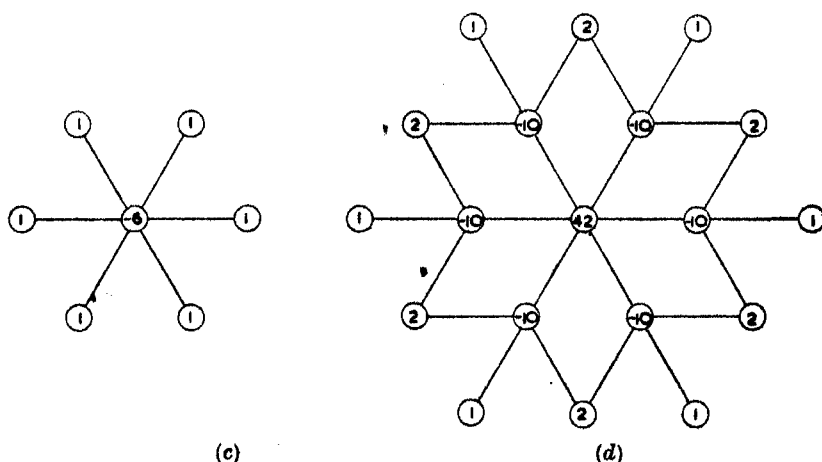
12. The 'patterns' on the right of figure 13 relate, similarly, to the finite-difference approximation to the biharmonic equation

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right]^2 w = Z(x, y), \quad (3)$$

in which $Z(x, y)$, as before, is specified. This equation governs (*inter alia*) the transverse displacement under pressure, also the 'Airy stress-function' χ which gives the stresses induced by extension, of a uniform elastic plate. I need not here develop a mechanical analogue of the finite-difference approximation: enough, that this leads (by similar reasoning) to the patterns shown.



'Relaxation patterns' for the operators ∇^2 and ∇^4 : square net ($N=4$)



'Relaxation patterns' for the operators ∇^2 and ∇^4 : triangular net ($N=6$)

FIGURE 13. 'Relaxation patterns' (Ref. 10, figures 5).

Our paper dealing with biharmonic analysis (Ref. 10) will not have open publication during the war, but I am allowed to show one solution—for the stresses in a standard 'cement briquette'. The problem is indicated in the left-hand diagram of figure 14: the 'jaws' apply a measured pull to the specimen, and calculation of stresses is needed to decide how far the assumption is justified, that uniform and simple tension is imposed across the waist. But here a question is presented as to the coefficient of friction between the jaws and the briquette: to leave it open, in calculation we have separated (as the photo-elastic method cannot do) the effects of (a) the longitudinal pull and (b) the transverse squeeze which the jaws exert.

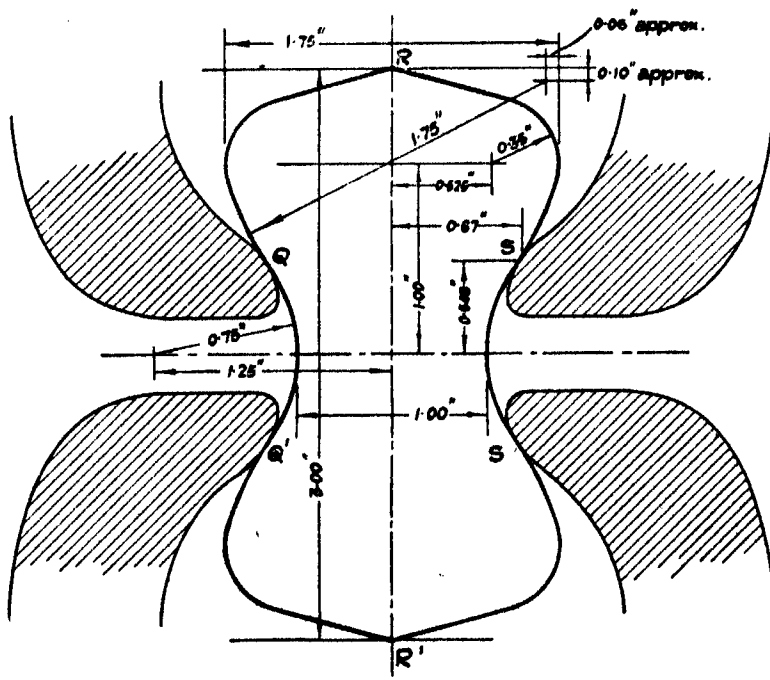
Our results for the second system (b) are shown in figures 14 and 15: figure 14 shows the cross-tension X_x , figure 15 shows on the left the longitudinal tension Y_y and on the right the shear stress-component X_y . It may be concluded that the squeeze (and hence the frictional coefficient) has little influence on the tension (Y_y) across the waist.

13. I leave the biharmonic equation (3) with the remark, that while the photo-elastic method will always have value as a means to *qualitative* understanding of a wanted stress-system, computation can now provide closer quantitative accuracy, and at no increased cost in respect of labour. It can, moreover, deal not only with problems in which edge tractions, but also with problems in which the edge displacements are specified, and even with 'mixed' boundary conditions (tractions specified at some points, displacements at others): Coker & Filon (1931, § 4.39) restrict the photo-elastic method to the first class of problem.

14. Before passing to other types of equation I must deal with another aspect of the relaxation process—whether applied to frameworks or to nets. *Is that process always convergent?*

This is a question for mathematicians, and I, from the standpoint of engineering science, would first propound another: *What is an 'exact' solution?* I have said that the orthodox procedure, if displacements are wanted, is to formulate equations in these as unknowns, with the specified loadings on their right-hand sides. But when we turn to reality from the rather artificial atmosphere of the examination hall, we are faced with the consideration that *in fact loadings are never known exactly*. To them, if to any of the data, 'tolerances' should be attached: 5 ± 0.1 tons here, 10 ± 0.3 tons there. But this is rarely done, even in cases confronted as problems of research; *because the doing of it would have no consequences for orthodox methods of attack*.

In our methods we fix attention, always, not on the wanted quantities (stresses or displacements) but on the data of the problem (on the loads). We assume these, always, to have some specified 'margin of uncertainty', and we 'whittle' away at them, accounting for more and more, until every 'residual' of a datum quantity has been brought within that margin. At that point we stop, asserting that to 'whittle' further would be not only waste of time, but *meaningless*. If you cannot tell the loads exactly



Part of figure 14; legend on facing page.

(and in practice you never can), then you have no claim to an exact solution. We need not apologize—we may even boast—that ours is 'mathematics with a fringe'.

15. It may not seem an important feature of the relaxation method, but in fact I believe it to be both fundamental and important philosophically, that in it we fix attention not on the wanted quantities but on the data of a problem. Ours is a tentative process, like the scraping process whereby, in engineering, a part is brought to coincide with surface-plate or gauge; and our 'margin of uncertainty' (in the loads) is akin to the margin of tolerance between a 'go' and a 'not go' gauge. 'Exact' measurement has

no meaning in metrology,—we can only work to some limiting accuracy; and so, I contend, exact solution has no meaning in a mathematics aimed at practical ends.

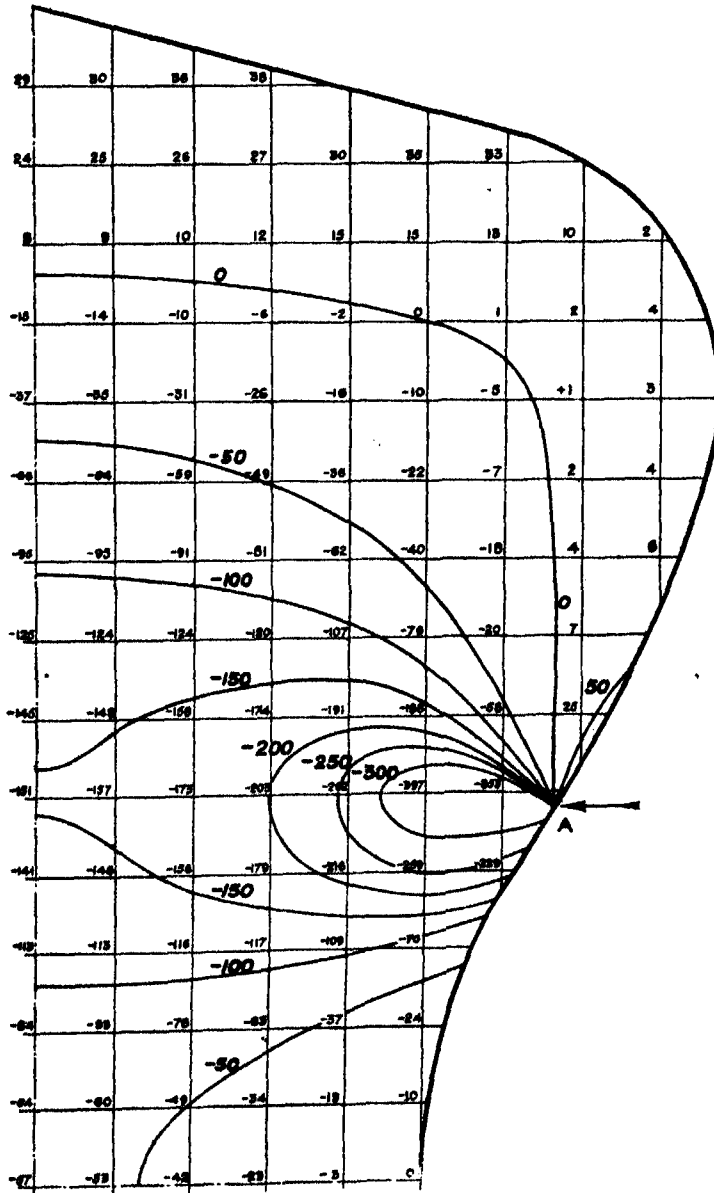
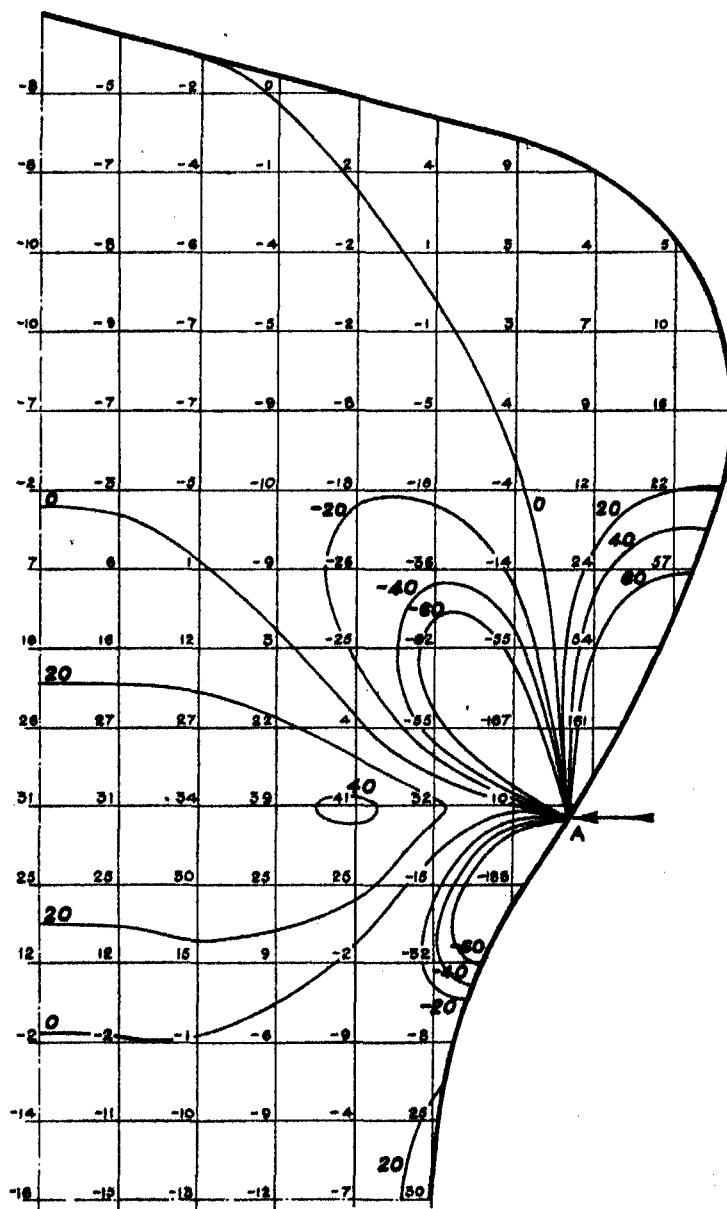


FIGURE 14. Concrete briquette problem, and computed values of X_* .
(Ref. 10, figures 21, 25a).

16. And now, as to the convergence of our computations for frameworks and the like. When we seek a stable equilibrium configuration for some mechanical system, we are (by a general theorem in mechanics) seeking



Part of figure 15; legend on facing page.

that configuration in which the total potential energy takes its minimum value; consequently any process which continually 'whittles at' (so as to reduce) the total potential energy must bring the configuration nearer

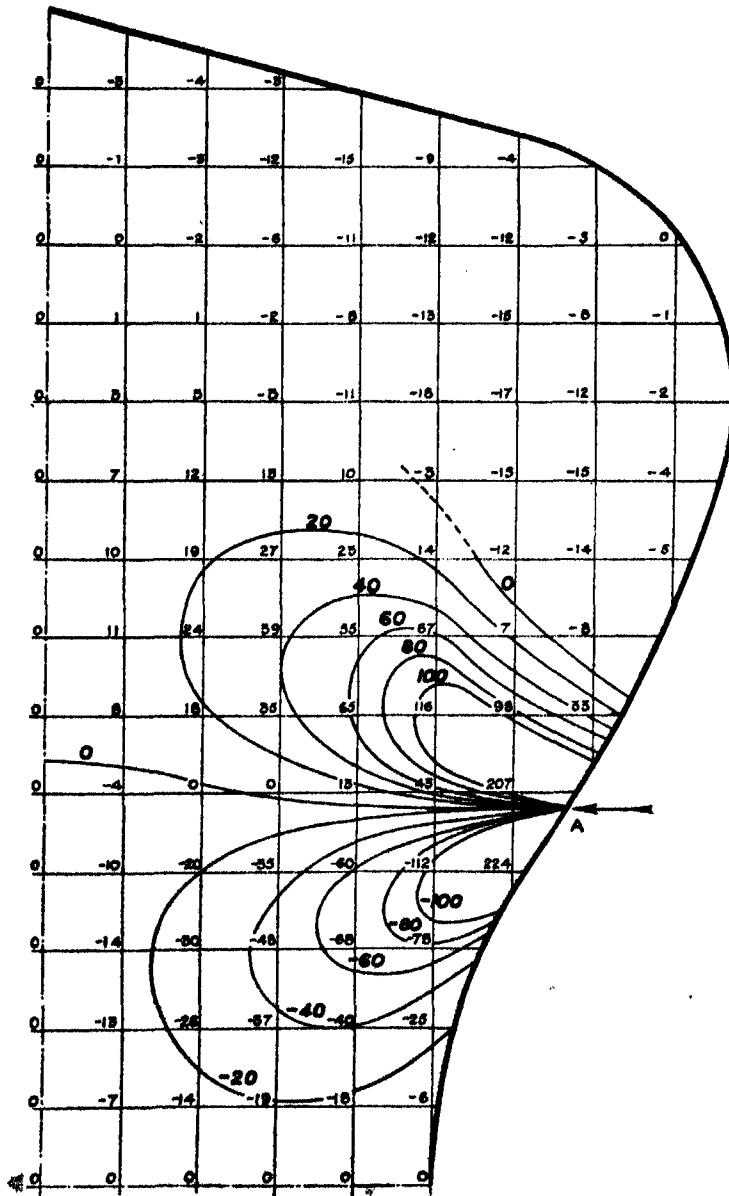
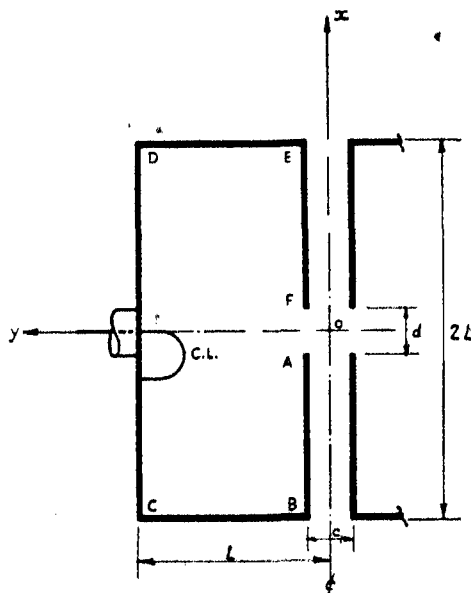


FIGURE 15. Concrete briquette problem: values of Y , and X ,
(Ref. 10, figures 25 b, c).

to what is wanted. An analogy has been suggested by Prof. Temple, which I quote for mathematicians to whom my picture of jacks and workmen makes a fainter appeal than to engineers:—In a valley devoid of friction the place at which a ball will come to rest is its lowest point: that point is the 'wanted configuration' in a system defined by two variables— x (east and west) and y (north and south). Proceeding by the rules of our relaxation method, we start from anywhere and proceed by 'stages': sometimes



Part of figure 16; legend on facing page.

east-and-west, sometimes north-and-south; but always proceeding downhill, and in every stage continuing until our path is level.* It is clear that we shall tend always towards our goal, unless the valley contains other 'stationary points' (hills) at which all paths are level; but this would mean that more than one point exists at which a ball can come to rest, and that possibility (in mechanical systems) can usually be excluded by a theorem of uniqueness of solution.

You will observe that in adopting Prof. Temple's analogy I have imposed no restriction except this on the shape of the valley,—that is, I have not assumed the total potential energy to be a quadratic function of the

* Slope is here the analogue of 'residual force'. Friction will give a 'margin of uncertainty'.

displacements: consequently my argument is *not* restricted to systems in which Hooke's law is satisfied,—it only requires that the wanted con-

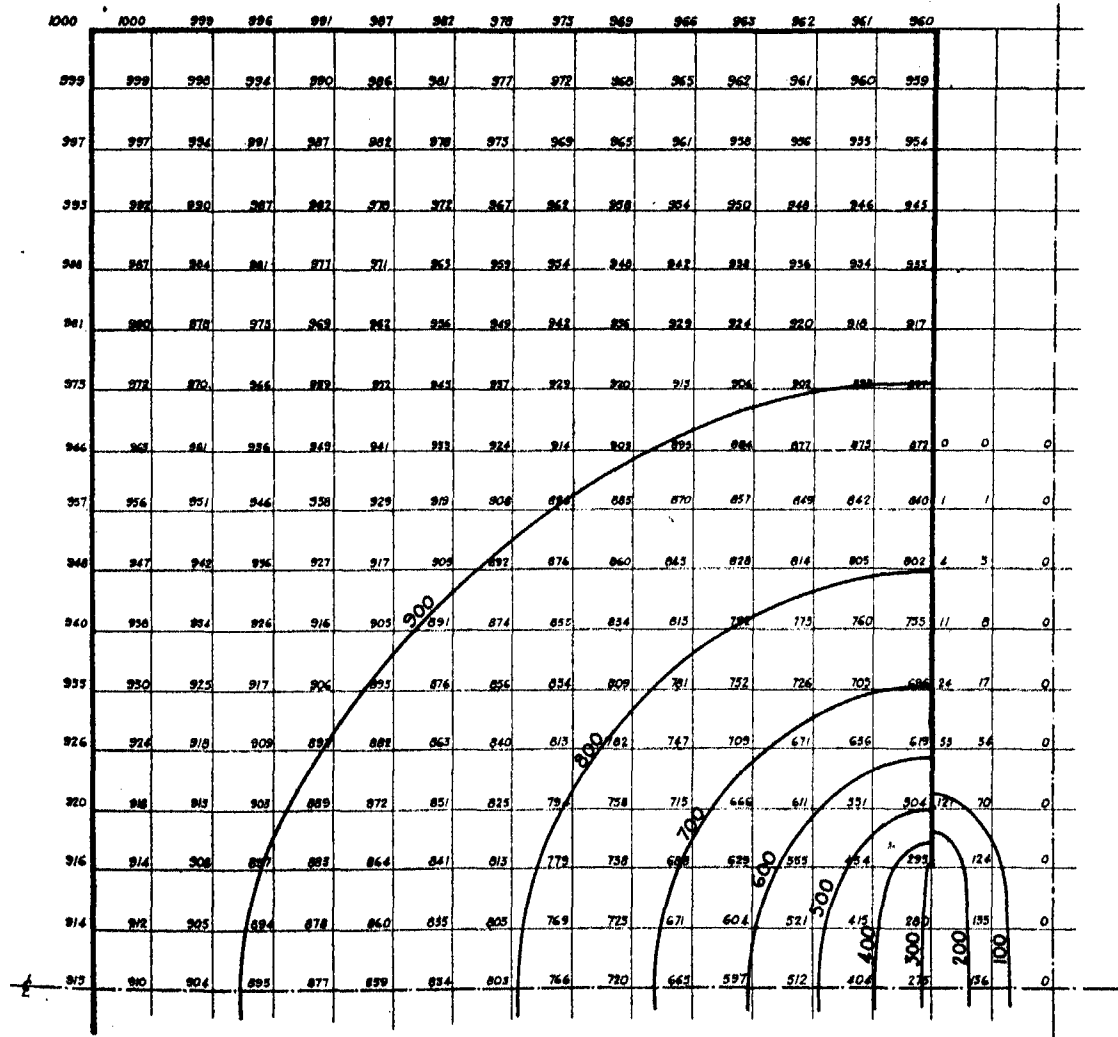


FIGURE 16. 'Wave guide' solution (Ref. 12, figures 3 and 4). The contours may be interpreted as relating to free transverse vibrations of an unloaded membrane.

figuration shall be unique. The point has importance, because we are using relaxation methods, now, on non-linear systems. In relating to linear systems (the main concern of my book) Temple has already undertaken

a mathematical defence of our methods: I now state my belief that a wider defence is possible.

17. I pass to a review of our latest work. First, not only statical, but *eigenwerte* (characteristic number) problems can be treated: I have given the underlying theory in my book, where it is exemplified by torsional oscillations in shafting and by the elastic stability of struts. The methods there described we have applied without essential change to problems in 'field physics', treated with a use of 'relaxation nets'; e.g. (1) to transverse vibrations of membranes and (2) to electro-magnetic oscillations in two dimensions, both of which problems entail a governing equation of the form

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right] w + \lambda w \cdot F(x, y) = 0, \quad (4)$$

with $F(x, y)$ specified; also to the elastic stability of plane frameworks or of flat plating to transverse 'buckling',—this last a problem of which the governing equation is

$$\lambda D \nabla^4 w + \frac{\partial}{\partial x} \left(P_x \frac{\partial w}{\partial x} - S \frac{\partial w}{\partial y} \right) + \frac{\partial}{\partial y} \left(P_y \frac{\partial w}{\partial y} - S \frac{\partial w}{\partial x} \right) = 0, \quad (5)$$

with P_x, S, P_y specified.* *Eigenwerte* solutions to (4) or (5) must yield not only the 'natural frequency' or 'critical loading' but also the associated mode of distortion. Figure 16 shows the computed mode for an electro-magnetic ('wave guide') problem; figures 17 and 18 explain a distinctly difficult problem in elastic stability.

18. Figure 17 shows a rectangular strip of plating, loaded (in its own plane) by forces uniformly distributed. The plotted curves are contours of the principal *compressive* stress, and are confined (as obviously they must be) to the bottom half. This system of stresses (chosen as being exactly calculable) is closely representative of what obtains in the web of a deep Ξ -girder or in the wing spar of an aeroplane; and in this connexion the question is presented, whether the web will be elastically stable. The tendency to buckle, since it is due to compressive stress, must increase sharply towards the right-hand bottom corner;† but close to the edges it is prevented by the clamping which is there presumed, so we may expect the mode of distortion to have its greatest amplitude somewhere near this corner, and elsewhere to involve no serious deflexions. This expectation is realized (figure 18).

* This problem was exemplified in the lecture by a model of a 'through girder' bridge.

† Cf. the values of stress-intensity which are attached to the contours in figure 17.

19. The mode exhibited in figure 18 was far from simple to compute, owing to the nature of the loading and of the boundary conditions: even

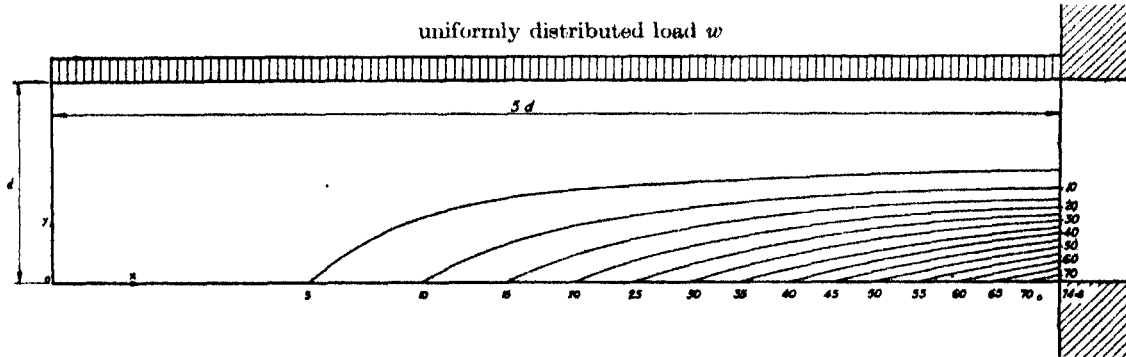


FIGURE 17. Loaded web of cantilever girder (Ref. 11, figure 11).
Contours exhibit the variation of the principal compressive stress.

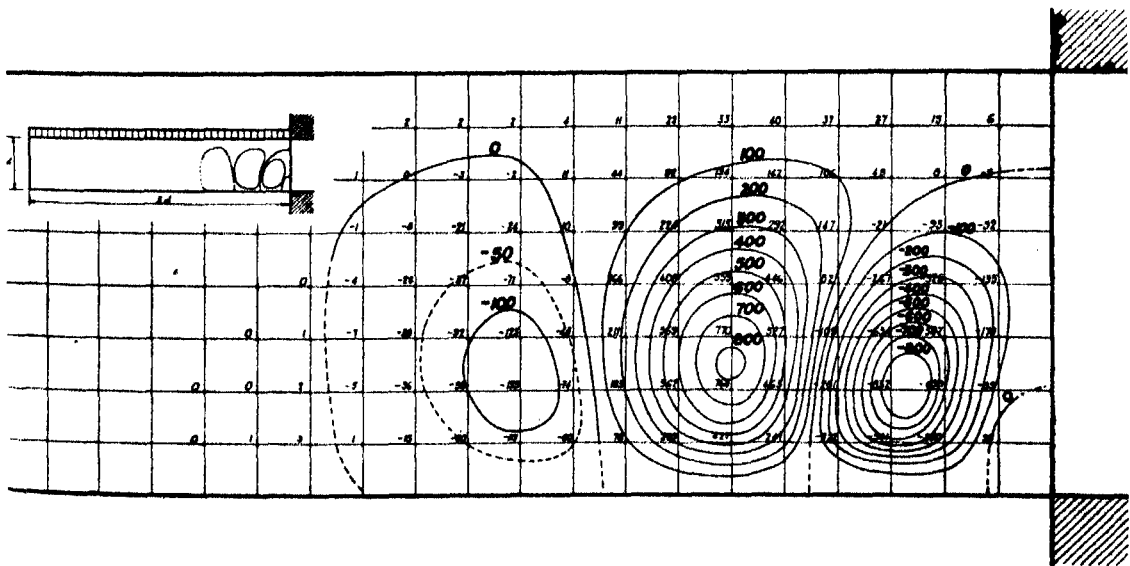


FIGURE 18. Mode of distortion under critical loading of cantilever girder web (Ref. 11; figure 12).

(The absolute magnitude of the distortion is indefinite.)

in simple cases the same is true when the mode is characterized by 'large deflexions'—'well-developed buckling'. H. Wagner has emphasized (1929) the importance for aeronautics of this latter class of problems. They are

very difficult, as may be judged from the governing equations which follow (v. Karman 1910):

$$\left. \begin{aligned} \nabla^4 \chi &= E \left\{ \left(\frac{\partial^2 w}{\partial x \partial y} \right)^2 - \frac{\partial^2 w}{\partial x^2} \frac{\partial^2 w}{\partial y^2} \right\}, \\ \nabla^4 w &= \frac{2h}{D} \left[\frac{Z}{2h} + \frac{\partial^2 \chi}{\partial y^2} \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 \chi}{\partial x^2} \frac{\partial^2 w}{\partial y^2} - 2 \frac{\partial^2 \chi}{\partial x \partial y} \frac{\partial^2 w}{\partial x \partial y} \right]. \end{aligned} \right\} \quad (6)$$

As you see, these are non-linear simultaneous equations relating two dependent variables, and the few solutions which are known entail very elaborate analysis. Recently we have found that—at considerable cost in labour—they too yield to relaxational attack. We wanted a test of accuracy, and for that reason have treated problems already solved: figures 19 and 20 show that in these our accuracy is more than sufficient, and again I would emphasize that we could have treated *any* shape of boundary.

Figure 19 shows, to a base (μ) proportional to the intensity of transverse loading, plottings of the central deflexion (w_0) and of three practically important stresses. I need not here go into details of the problem: what

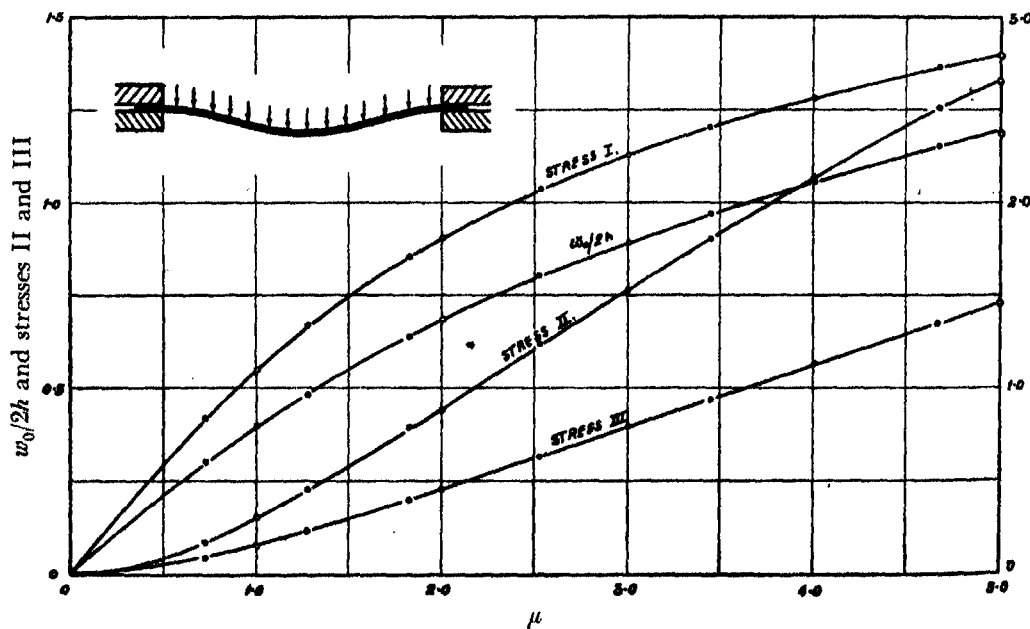


FIGURE 19. Transversely-loaded circular plate: comparison with exact solution by S. Way (Ref. 15, figure 4).

matters now is the order of the agreement between our results (the open circles) and Way's (the points shown black).

20. Clearly the agreement is satisfactory, and equally good agreement is revealed, in figure 20, between our results and those of Friedrichs and Stoker for a circular plate sustaining edge thrust. (The comparison is between our results—the broken curve—and the open circles.) A word should be said regarding the parabolas lettered *A, B, C, D, E*, because the method used to obtain them (an extension of 'Rayleigh's principle') seems likely to have great value for these difficult problems of 'well-developed buckling'. Having the 'critical load for small deflexions' ($\lambda = 1.1$ in figure 20), together with the associated mode, without additional computation beyond a single integration we were able to construct the parabola *A*, and we could show that *the wanted (broken-line) curve must lie below this*. Consequently a relatively short exploration (indicated by the points numbered 1, 2, 3, 4, 5 on the relevant ordinate) led to the required solution for $\mu = 2$; having this we could similarly construct the parabola *C*, thus shortening the exploration for $\mu = 3$; and so on. The resulting (dotted) curve extends well beyond the elastic range of ordinary material.

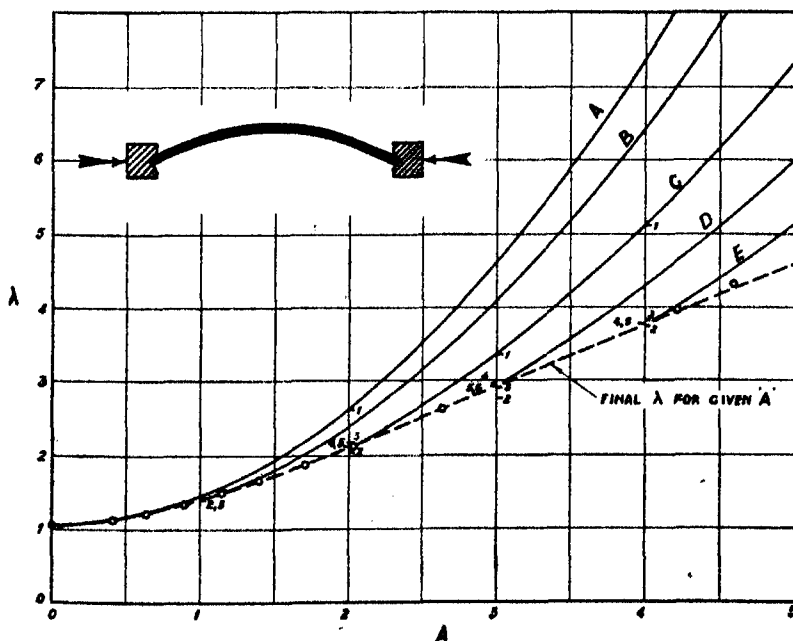


FIGURE 20. Elastic stability of compressed circular plate; comparison with exact solution of Friedrichs and Stoker (Ref. 15, figure 8).

21. Also complicated, although now the governing equations are linear, are problems of stress-calculation for solids of revolution. We have dealt with several in a paper recently communicated, and I show results for the hardest of them—stress-determination for a toroidal hook. Here, strictly treated, not one of the six stress-components is zero; all can be expressed in terms of two related functions ϕ and ψ , and the boundary conditions are *three* in number, therefore hard to satisfy.

Stress-expressions in the hook problem (Ref. 20, equations (61), (75), (76)):

$$\left. \begin{aligned} r^2 \cdot \widehat{z z} &= \frac{\partial \phi}{\partial r}, \\ r^2 \cdot \widehat{r r} &= \frac{\partial}{\partial z} \left(\frac{1}{r} \frac{\partial \psi}{\partial r} - \phi \right), \\ r^2 \cdot \widehat{\theta z} &= \frac{\partial}{\partial z} \left(\left[\frac{\partial^2}{\partial r^2} - \frac{2}{r} \frac{\partial}{\partial r} \right] \psi + \phi \right), \end{aligned} \right\} \quad (7)$$

$$\left. \begin{aligned} r^3 \cdot \widehat{r r} &= Q(r^2 + \sigma z^2) - \sigma A(zr^2 + \frac{1}{3}z^3) - \left[r \frac{\partial}{\partial r} - 3 \right] \phi - 2 \frac{\partial^2 \psi}{\partial z^2} \\ &\quad - \left[r \frac{\partial}{\partial r} - (3 - \sigma) \right] \vartheta \frac{2}{3} \psi, \\ r^3 \cdot \widehat{\theta \theta} &= -Q(r^2 + \sigma z^2) - A\{(2 + \sigma) zr^2 - \frac{1}{3}\sigma z^3\} - 3\phi + 2 \frac{\partial^2 \psi}{\partial z^2} \\ &\quad - \left[\sigma r \frac{\partial}{\partial r} + (3 - \sigma) \right] \vartheta \frac{2}{3} \psi, \\ -r^3 \cdot \widehat{r \theta} &= Q(r^2 + \sigma z^2) - \sigma A(zr^2 + \frac{1}{3}z^3) - \left[r \frac{\partial}{\partial r} - 3 \right] \phi \\ &\quad + \left[r \frac{\partial}{\partial r} - 2 \right] \frac{\partial^2 \psi}{\partial z^2} - \left[r \frac{\partial}{\partial r} - (3 - \sigma) \right] \vartheta \frac{2}{3} \psi. \end{aligned} \right\} \quad (8)$$

Boundary conditions:

$$\left. \begin{aligned} \phi &= \frac{\partial^2 \psi}{\partial z \partial r} = 0, \\ \cos(r, \nu) \left\{ Q(r^2 + \sigma z^2) - \sigma A(zr^2 + \frac{1}{3}z^3) - 2 \frac{\partial^2 \psi}{\partial z^2} \right. \\ &\quad \left. - \left[r \frac{\partial}{\partial r} - (3 - \sigma) \right] \vartheta \frac{2}{3} \psi \right\} = r \frac{\partial \phi}{\partial \nu}. \end{aligned} \right\} \quad (9)$$

Equations (7)–(9) present the mathematical problem, which looks distinctly ‘fierce’: by contrast, figure 21 shows how simple may be the implications of a complex mathematical solution. The contours of tensile

stress $\theta\theta$, for the heavily loaded section AB , are not straight as assumed in the customary theory, but they are very flat curves. Figure 21 relates to hooks of standard B.S.I. sections, and is based on a paper recently communicated (Ref. 18).

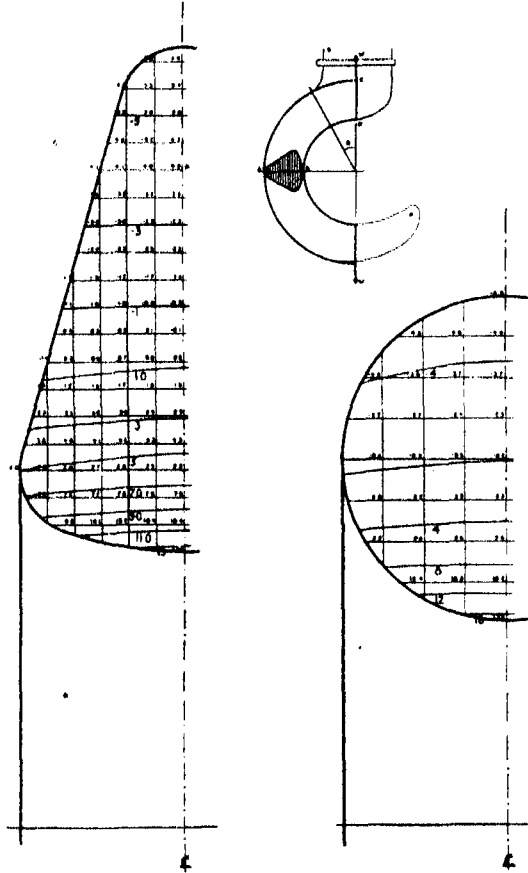


FIGURE 21. Tensile stress-distribution in two B.S.I. standard hooks (Ref. 18, figures 3, 4).

22. This work on large deflexions and on hooks will serve to illustrate yet another advantage of the relaxation method. I have shown by my examples that its errors are made apparent at every stage; I have emphasized its tentative quality, which has enabled us to overcome the difficulties of boundaries not defined initially, and of data characterized by discontinuities like the stress-strain diagram of a plastic material; and I might have mentioned in addition, that it can utilize just as easily experi-

mental data, presented in curves, as data expressible mathematically—e.g. a pressure-density relation found experimentally for steam. My last point is that by its freedom from restriction, and by its tentative quality, it lends itself ideally to *iterative* methods of solution. This, more and more, is a feature we are tending to exploit as we enter the field of *non-linear equations*.

In fact, when he discards 'orthodox' for 'relaxational' methods an investigator finds his problem quite transformed,—the incidence of computational difficulties has completely shifted. I do not mean that they have disappeared: full scope remains for inventive artifice. But he can take for granted (as involving nothing more than time and labour) operations that formerly he would have regarded as beyond his power—e.g. potential solutions for a series of slightly differing boundaries. His tools become more powerful, though admittedly less exact.

23. I have not time to elaborate this point, but I can exemplify it by recent solutions in some fields as yet hardly explored. One such field of study relates to *compressible* fluids, another to the strains in *plastic* material. I have touched on both already, and figures 22 and 23 show some of our latest results.

24. My last point relates to our presentation of results. What have we when we terminate our computations, deeming their accuracy sufficient?

No formal solution, merely a 'grid' of computed values, coarse or fine as our effort has been brief or sustained. But this we must derive from any formal solution before it can be put to practical use: the pure analyst, when having given us a formal solution he goes home to his tea, from our standpoint has left his work half done. To any who object that it is not the whole picture, I reply that it is what we accept without question in a 'half-tone' reproduction (cf. figure 24).

Mr A. N. Black, of the Engineering Department at Oxford, has pursued this line of thought somewhat further, using a microscope to analyse part of a half-tone reproduction taken from a newspaper. Each elemental square was viewed in turn, and an estimate made and recorded of the fraction of its area which was 'black' (i.e. inked); then, from the record of these estimates, a diagram was constructed on squared paper to have the same fractions inked of its corresponding squares. No particular convention was adopted,—the inked fraction might have any shape or position in its square; but viewed from a suitable distance the resulting diagrams (figure 25) reproduce with striking fidelity the general quality of the original photograph. Few 'wanted functions' are likely to have more elaborate features!

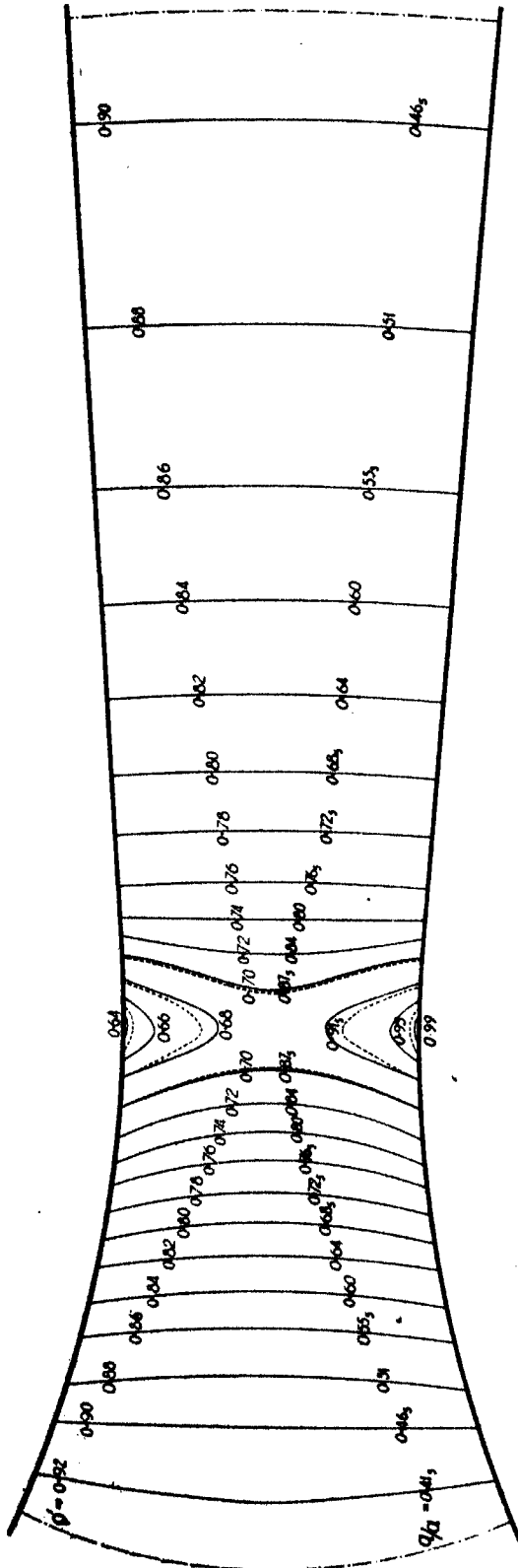


FIGURE 22. High-speed flow of gas through a nozzle: contours of density and velocity (Ref. 16, figure 11).

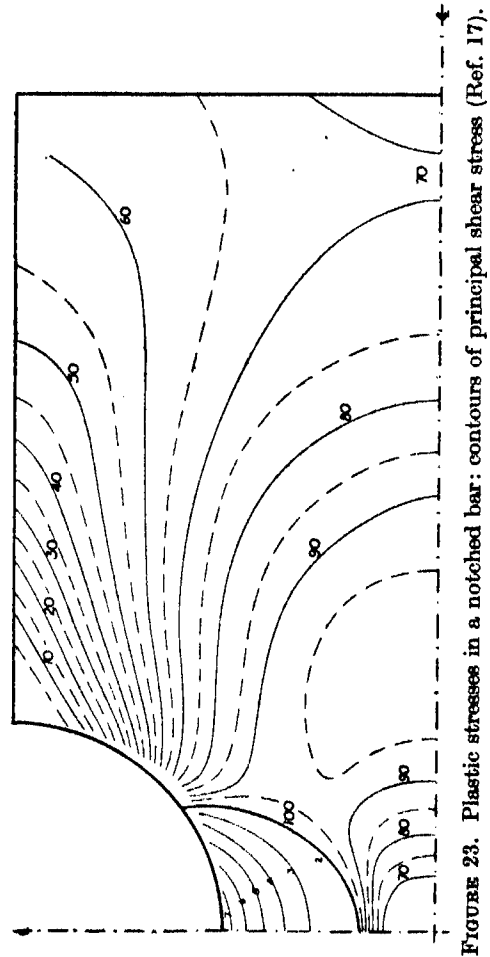


FIGURE 23. Plastic stresses in a notched bar: contours of principal shear stress (Ref. 17).

25. I have been compelled this afternoon to hurry, but I have tried to avoid detail and to give some impression of the kind of thing that can now



FIGURE 24. Photographic enlargement of 'half-tone' reproduction (by Mr V. Belfield).

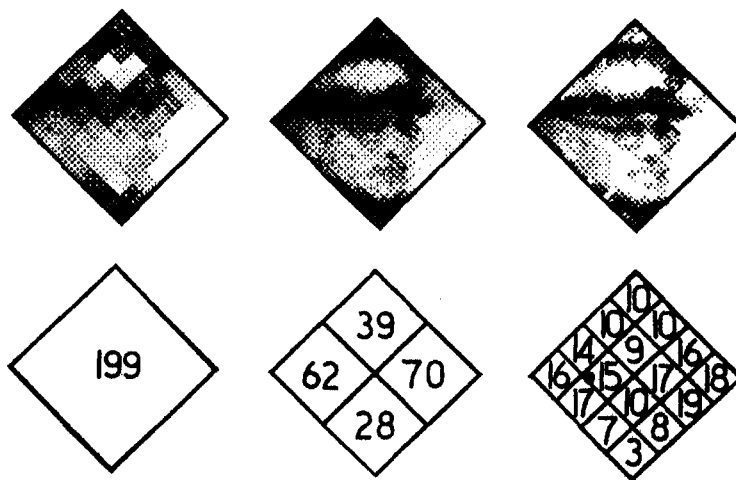


FIGURE 25

be done. It has been pleasant, in the past six years, to try our methods out against problems of steadily increasing difficulty, and to have met no failure so far, though the labour is sometimes heavy.

To exploit them fully—as is my natural desire—a problem should be ‘broken down’, like any problem of engineering construction, into jobs appropriate to varying degrees of skill. A novice soon acquires power to solve plane-potential problems; with longer experience other types of problem become familiar; and so on, up to the level at which real acumen is demanded for the planning of new research. This of course implies considerable facilities—a mathematical laboratory, or institute, of a kind which has more appeal in America than in this country; but it would be vain to hope for such developments in war-time, and instead I have tried to extend our *range* to the utmost, even at the cost of leaving attractive byways unexplored.

What once we regarded as an alternative method having some practical advantages, in our recent studies we are finding an indispensable weapon of research. I hope that some of my audience, who have heard me with such patience this afternoon, may perhaps see ways in which our methods may help them too.

REFERENCES

Ref.
no.

‘Stress-Calculation in Frameworks by the “Method of Systematic Relaxation of Constraints”.’

1. Parts I and II (Southwell 1935*a*): *Proc. Roy. Soc. A*, **151**, 56–95.
 2. Part III (Southwell 1935*b*): *Proc. Roy. Soc. A*, **153**, 41–76.
- ‘Relaxation Methods applied to Engineering Problems’:
3. I (Bradfield & Southwell 1937): ‘The deflexion of beams under transverse loading.’ *Proc. Roy. Soc. A*, **161**, 155–181.
 4. II (Black & Southwell 1938): ‘Basic theory, with application to surveying and to electrical networks, and an extension to gyrostatic systems.’ *Proc. Roy. Soc. A*, **164**, 447–467.
 5. III (Christopherson & Southwell 1938): ‘Problems involving two independent invariables.’ *Proc. Roy. Soc. A*, **168**, 317–350.
 6. IV (Bradfield, Christopherson & Southwell 1939): ‘Problems relating to elastic stability and vibrations.’ *Proc. Roy. Soc. A*, **169**, 289–317.
 7. V (Gandy & Southwell 1940): ‘Conformal transformation of a region in plane space.’ *Phil. Trans. A*, **238**, 453–475.
 8. VI (Pellew & Southwell 1940): ‘The natural frequencies of systems having restricted freedom.’ *Proc. Roy. Soc. A*, **175**, 262–290.
 9. VII (Shaw & Southwell 1941): ‘Problems relating to the percolation of fluids through porous materials.’ *Proc. Roy. Soc. A*, **178**, 1–17.
 10. VIIA (Fox & Southwell 1945): ‘Biharmonic analysis as applied to the flexure and extension of flat elastic plates.’ *Phil. Trans. A*, **239**, 419–460.

11. VII_B (Christopherson, Fox, Green, Shaw & Southwell 1945): 'The elastic stability of plane frameworks and of flat plating.' *Phil. Trans. A*, **239**, 461-487.
12. VII_C (Allen, Fox, Motz & Southwell 1945): 'Free transverse vibrations of membranes, with an application (by analogy) to two-dimensional oscillations in an electro-magnetic system.' *Phil. Trans. A*, **239**, 489-501.
13. VII_D (Allen, Fox & Southwell 1945): 'Stress distributions in elastic solids of revolution.' *Phil. Trans. A*, **239**, 503-539.
14. VIII (Southwell & Vaisey† 1943): 'Plane-potential problems involving specified normal gradients.' *Proc. Roy. Soc. A*, **182**, 129-151.
15. VIII_A (Green & Southwell 1945): 'Problems relating to large transverse displacements of thin elastic plates.' *Phil. Trans. A*, **239**, 541-580.
16. IX (Green & Southwell 1944): 'High-speed flow of compressible fluid through a two-dimensional nozzle.' *Phil. Trans. A*, **239**, 367-386.
17. *XII_A (Allen & Southwell 194-): 'Plastic strains in flat plates and solids of revolution.'
18. †'On the Stresses in Hooks, and their determination by Relaxation Methods' (Fox & Southwell 194-). Accepted for publication by Inst. Mech. Engrs.
19. *Relaxation Methods in Engineering Science* (Southwell 1940). Oxford Univ. Press.
20. 'Some practically important stress-systems in solids of revolution' (Southwell 1942). *Proc. Roy. Soc. A*, **180**, 367-396.

Casagrande, A. 1937 *J. New Eng. Wat. Wks Ass.* **51**, 131-173.

Coker, E. G. & Filon, L. N. G. 1931 *A treatise on photo-elasticity*. Camb. Univ. Press.

Davison, B. & Rosenhead, L. 1940 *Proc. Roy. Soc. A*, **175**, 346-365.

Griffith, A. A. & Taylor, G. I. 1917 *Proc. Instn. Mech. Engrs*, 755-789.

Hele-Shaw, H. S. & Hay, A. 1900 *Phil. Trans. A*, **195**, 303-327.

Karman, T. v. 1910 *Encyc. math. Wiss.* **4** (iv), 349.

Nadai, A. 1931 *Plasticity*. McGraw Hill Co.

Prandtl, L. 1903 *Phys. Z.* **4**, 758-759.

Prandtl, L. 1923 *Z. angew. Math. Mech.* **3**, no. 6, 442-446.

Wagner, H. 1929 *Z. Flugtech.* **20**, 200-207.

* Not yet communicated (July 1945).

† Not yet published (July 1945).

Stress systems in aeolotropic plates. VI

By A. E. GREEN

(Communicated by G. I. Taylor, F.R.S.—Received 26 August 1943)

The distribution of stress due to hyperbolic notches in an aeolotropic plank under tension or flexure is discussed theoretically when the plank has two directions of symmetry at right angles. Numerical work is carried out using the elastic constants found in experiments made with specimens cut in the L, T plane from spruce wood which is highly aeolotropic. The calculated stresses are used in conjunction with measurements of ultimate strength to determine the kind of failure which might be expected near notches in a spruce plank under tension.

INTRODUCTION

1. Most of our information about the effect of notches or cuts made in the edges of an isotropic plate under tension or flexure has been obtained from photo-elastic experiments (see Coker & Filon 1931). The only problems of this nature which so far seem to have received exact theoretical treatment are those of a semi-infinite tension member bounded by a straight edge which contains a semi-circular notch (Maunsell 1936), and a plate of very large width with hyperbolic grooves, the plate being under tension or flexure (Neuber 1933, 1934, 1935).^{*} The theoretical solution for the latter problem is very simple, and although the problem is slightly artificial it throws considerable light from the theoretical point of view on the nature of the effect of notches on stress distributions.

It seems that any attempt to give an exact theoretical account of the effect of notches in the edges of stressed aeolotropic plates is likely to meet with considerable difficulty. In previous papers, however (see references), it has been found that the stresses in aeolotropic plates may differ radically from those in isotropic plates, so that the application of isotropic results to aeolotropic materials may be very misleading and may in some cases give entirely wrong results. Since at present no experimental method which is comparable to that used in photo-elastic experiments is available for aeolotropic bodies, it is important that as many theoretical results as possible should be obtained for aeolotropic materials, especially for the many types of wood which are used in practice.

^{*} Professor Taylor has pointed out that stresses in an isotropic plate containing hyperbolic notches were first considered by Dr A. A. Griffith, *Rep. Memor. Aero. Res. Comm., Lond.*, 1928, no. 1152.

It has been found that the problems of the stresses due to hyperbolic notches in a very wide plate under tension or flexure have simple solutions for aeolotropic as well as for isotropic materials, and the solutions are given in the present paper.

In Paper III of this series the stresses in a spruce tension member containing a circular hole were found theoretically, and the results were applied in conjunction with measurements of ultimate yield stresses for spruce in order to determine the kind of failure which might be expected near a hole in a highly stressed spruce plank. In the present paper theoretical calculations are made for the same specimen of spruce as before, and these results are also combined with the measurements of ultimate yield stresses in order to estimate the kind of failure which might be expected near notches in a spruce plank under tension.

GENERAL STRESS FUNCTIONS

2. Consider an aeolotropic plate which has two directions of symmetry parallel to the co-ordinate axes (x, y) . The z -plane, which is the plane of the plate, is transformed to a suitable region in the ζ ($= \xi + i\eta$) plane by the relation

$$z = F(\zeta). \quad (2.1)$$

Also
$$\frac{dz}{d\zeta} = F'(\zeta) = Je^{i\phi}, \quad (2.2)$$

where ϕ is the angle between the tangent to the curve $\eta = \text{constant}$ through any point, and the x -axis. It has been shown in a previous paper (Green 1945) that the stresses in rectangular co-ordinates can be taken to be the real parts of

$$\left. \begin{aligned} \widehat{xx} &= -(1 - \gamma_1)^2 f''(z + \gamma_1 \bar{z}) - (1 - \gamma_2)^2 g''(z + \gamma_2 \bar{z}), \\ \widehat{yy} &= (1 + \gamma_1)^2 f''(z + \gamma_1 \bar{z}) + (1 + \gamma_2)^2 g''(z + \gamma_2 \bar{z}), \\ \widehat{xy} &= -i(1 - \gamma_1^2) f''(z + \gamma_1 \bar{z}) - i(1 - \gamma_2^2) g''(z + \gamma_2 \bar{z}). \end{aligned} \right\} \quad (2.3)$$

The stresses in the orthogonal co-ordinates (ξ, η) are related to the stresses in the cartesian co-ordinates (x, y) by the equations

$$\left. \begin{aligned} \widehat{\xi\xi} &= \widehat{xx} \cos^2 \phi + \widehat{xy} \sin 2\phi + \widehat{yy} \sin^2 \phi, \\ \widehat{\eta\eta} &= \widehat{xx} \sin^2 \phi - \widehat{xy} \sin 2\phi + \widehat{yy} \cos^2 \phi, \\ \widehat{\xi\eta} &= \frac{1}{2}(\widehat{yy} - \widehat{xx}) \sin 2\phi + \widehat{xy} \cos 2\phi. \end{aligned} \right\} \quad (2.4)$$

Thus, using (2.2), (2.3) and (2.4),

$$\left. \begin{aligned} \xi\xi &= -\frac{\{F'(\xi) - \gamma_1 \bar{F}'(\bar{\xi})\}^2}{F'(\xi) \bar{F}'(\bar{\xi})} f''(z + \gamma_1 \bar{z}) - \frac{\{F'(\xi) - \gamma_2 \bar{F}'(\bar{\xi})\}^2}{F'(\xi) \bar{F}'(\bar{\xi})} g''(z + \gamma_2 \bar{z}), \\ \eta\eta &= \frac{\{F'(\xi) + \gamma_1 \bar{F}'(\bar{\xi})\}^2}{F'(\xi) \bar{F}'(\bar{\xi})} f''(z + \gamma_1 \bar{z}) + \frac{\{F'(\xi) + \gamma_2 \bar{F}'(\bar{\xi})\}^2}{F'(\xi) \bar{F}'(\bar{\xi})} g''(z + \gamma_2 \bar{z}), \\ \xi\eta &= -i \left\{ \frac{F'(\xi)}{\bar{F}'(\bar{\xi})} - \frac{\gamma_1 \bar{F}'(\bar{\xi})}{F'(\xi)} \right\} f''(z + \gamma_1 \bar{z}) - i \left\{ \frac{F'(\xi)}{\bar{F}'(\bar{\xi})} - \frac{\gamma_2 \bar{F}'(\bar{\xi})}{F'(\xi)} \right\} g''(z + \gamma_2 \bar{z}), \end{aligned} \right\} \quad (2.5)$$

where the real parts only of these expressions are to be used. For some calculations it is useful to observe that $\xi\xi + \eta\eta$ is given by the real part of

$$\xi\xi + \eta\eta = 4\gamma_1 f''(z + \gamma_1 \bar{z}) + 4\gamma_2 g''(z + \gamma_2 \bar{z}). \quad (2.6)$$

The region in the z -plane is supposed to be bounded by two branches of a hyperbola (see figure 3), and a suitable transformation for such a region is

$$z = c \sinh \xi, \quad (2.7)$$

where $\eta = \pm \alpha$ represent the hyperbolic boundaries. The x -axis is along the line of the arrows in figure 3. If 2α is the width of the plate at the notches and if ρ is the radius of curvature of the hyperbola at the bottom of a notch, then

$$a = c \sin \alpha, \quad a/\rho = \tan^2 \alpha. \quad (2.8)$$

From (2.2) and (2.7) it is found that on the boundary of the plate

$$\left. \begin{aligned} \cos^2 \phi &= \frac{2 \cosh^2 \xi \cos^2 \alpha}{\cosh 2\xi + \cos 2\alpha}, & \sin^2 \phi &= \frac{2 \sinh^2 \xi \sin^2 \alpha}{\cosh 2\xi + \cos 2\alpha}, \\ \sin 2\phi &= \frac{\sinh 2\xi \sin 2\alpha}{\cosh 2\xi + \cos 2\alpha}. \end{aligned} \right\} \quad (2.9)$$

THE TENSION PROBLEM

3. Consider the stress system which can be derived from the expressions

$$\left. \begin{aligned} (\gamma_1 - \gamma_2) f''(z + \gamma_1 \bar{z}) &= A(1 + \gamma_2) \{c^2(1 + 2\gamma_1 \cos 2\alpha + \gamma_1^2) + (z + \gamma_1 \bar{z})^2\}^{-\frac{1}{2}}, \\ (\gamma_2 - \gamma_1) g''(z + \gamma_2 \bar{z}) &= A(1 + \gamma_1) \{c^2(1 + 2\gamma_2 \cos 2\alpha + \gamma_2^2) + (z + \gamma_2 \bar{z})^2\}^{-\frac{1}{2}}. \end{aligned} \right\} \quad (3.1)$$

These give a total force X in the x -direction across any section $x = \text{constant}$ of the plate, where

$$A\{(1 + \gamma_1)(1 - \gamma_2)\phi_2 - (1 + \gamma_2)(1 - \gamma_1)\phi_1\} = aT(\gamma_1 - \gamma_2) = \frac{1}{2}X(\gamma_1 - \gamma_2), \quad (3.2)$$

T being the average stress across the smallest section, and where

$$\tan \phi_1 = \frac{1-\gamma_1}{1+\gamma_1} \tan \alpha, \quad \tan \phi_2 = \frac{1-\gamma_2}{1+\gamma_2} \tan \alpha. \quad (3.3)$$

From (2.5) and (3.1) it is found that the boundaries $\eta = \pm \alpha$ are free from applied stress, that is, $\xi \bar{\eta}_e = 0$ and $\bar{\eta} \eta_e = 0$. Also, from (2.6) and (3.1), it is found that the edge stress takes the form

$$\begin{aligned} \xi \bar{\xi}_e \{ (1+\gamma_1^2) (\cosh 2\xi + \cos 2\alpha) + 2\gamma_1 (1 + \cosh 2\xi \cos 2\alpha) \} \\ \times \{ (1+\gamma_2^2) (\cosh 2\xi + \cos 2\alpha) + 2\gamma_2 (1 + \cosh 2\xi \cos 2\alpha) \} \\ = \lambda T \cosh \xi (\cosh 2\xi + \cos 2\alpha), \end{aligned} \quad (3.4)$$

$$\begin{aligned} \text{where} \quad \lambda &= 4A(1-\gamma_1\gamma_2)(1+\gamma_1)(1+\gamma_2) \sin 2\alpha / aT \\ &= \frac{4(\gamma_1-\gamma_2)(1-\gamma_1\gamma_2)(1+\gamma_1)(1+\gamma_2) \sin 2\alpha}{(1+\gamma_1)(1-\gamma_2)\phi_2 - (1+\gamma_2)(1-\gamma_1)\phi_1}. \end{aligned} \quad (3.5)$$

Stresses for an isotropic material, which have been obtained by Neuber (1933), may be deduced from the above formulae by finding their limit as $\gamma_1 \rightarrow \gamma_2 \rightarrow 0$. In particular, it is found that

$$\xi \bar{\xi}_e = \frac{\lambda T \cosh \xi}{\cosh 2\xi + \cos 2\alpha}, \quad \lambda = \frac{4A \sin 2\alpha}{aT} = \frac{4 \tan \alpha}{\tan \alpha + \alpha(1 + \tan^2 \alpha)}. \quad (3.6)$$

THE FLEXURE PROBLEM

4. The stress system to be considered in this section is obtained from the functions

$$\begin{aligned} (\gamma_1 - \gamma_2) f''(z + \gamma_1 \bar{z}) &= iB(1 - \gamma_2^2)(z + \gamma_1 \bar{z}) \{ c^2(1 + 2\gamma_1 \cos 2\alpha + \gamma_1^2) + (z + \gamma_1 \bar{z})^2 \}^{-\frac{1}{2}}, \\ (\gamma_2 - \gamma_1) g''(z + \gamma_2 \bar{z}) &= iB(1 - \gamma_1^2)(z + \gamma_2 \bar{z}) \{ c^2(1 + 2\gamma_2 \cos 2\alpha + \gamma_2^2) + (z + \gamma_2 \bar{z})^2 \}^{-\frac{1}{2}}. \end{aligned} \quad (4.1)$$

These stresses give a couple M in the plane of the plate, across any section $x = \text{constant}$, where

$$M = \frac{Ba^2(1-\gamma_1)(1-\gamma_2)}{\gamma_1-\gamma_2} \left\{ \frac{(1-\gamma_1)(1+\gamma_2)\phi_1}{\sin^2 \phi_1} - \frac{(1-\gamma_2)(1+\gamma_1)\phi_2}{\sin^2 \phi_2} \right\}. \quad (4.2)$$

On the boundaries $\eta = \pm \alpha$ it is found that $\xi \bar{\eta}_e = 0$ and $\bar{\eta} \eta_e = 0$ and

$$\begin{aligned} \xi \bar{\xi}_e \{ (1+\gamma_1^2) (\cosh 2\xi + \cos 2\alpha) + 2\gamma_1 (1 + \cosh 2\xi \cos 2\alpha) \} \\ \times \{ (1+\gamma_2^2) (\cosh 2\xi + \cos 2\alpha) + 2\gamma_2 (1 + \cosh 2\xi \cos 2\alpha) \} \\ = \mp aM (\cosh 2\xi + \cos 2\alpha) / a^2, \end{aligned} \quad (4.3)$$

where

$$\begin{aligned}\mu &= 4Ba^2(1-\gamma_1\gamma_2)(1-\gamma_1^2)(1-\gamma_2^2)\sin 2\alpha/M \\ &= \frac{4(\gamma_1-\gamma_2)(1-\gamma_1\gamma_2)(1+\gamma_1)(1+\gamma_2)\sin 2\alpha}{\frac{(1-\gamma_1)(1+\gamma_2)\phi_1}{\sin^2\phi_1} - \frac{(1-\gamma_2)(1+\gamma_1)\phi_2}{\sin^2\phi_2}}.\end{aligned}\quad (4.4)$$

Stresses for an isotropic material, which have been found by Neuber (1934, 1935), may be deduced by a limiting process. In particular it is found that

$$\xi_{\xi\xi} = \mp \frac{\mu M}{(\cosh 2\xi + \cos 2\alpha)a^2}, \quad \mu = \frac{4Ba^2\sin 2\alpha}{M} = \frac{2\sin 2\alpha \tan^2\alpha}{\tan\alpha + \alpha(\tan^2\alpha - 1)}.\quad (4.5)$$

NUMERICAL DISCUSSION: THE TENSION PROBLEM

5. Calculations have been carried out for a specimen of spruce wood whose elastic constants are shown in table 1:

TABLE 1

	s_{11}	s_{22}	s_{12}	s_{66}	α_1	α_2	γ_1	γ_2
spruce	15.5	0.587	-0.33	11.5	16.91	1.56	0.608	0.111

With these values of the constants the grain of the wood is parallel to the y -axis. If the values of s_{11} and s_{22} are interchanged (which means a change in sign of γ_1 and γ_2), then the grain is parallel to the x -axis. The inverses of the constants s_{11}, \dots, s_{66} are measured in 10^3 kg./sq. mm. The calculations have been confined to a plate containing notches which correspond to a value of α equal to 45° , so that $\rho = a$. The values of the stresses along the edges of the plate have been evaluated by using formulae (2.4), (2.9), (3.4) and (3.5), and the results are given in table 2. The cross-stress $\hat{x}\hat{x}$ at the smallest section of the beam, which may be obtained from (2.3) and (3.1), has also been evaluated and the results are reproduced in table 3, together with the values of $\hat{y}\hat{y}$.

The distribution of $\xi\xi$ over one-quarter of the edge of the plate is shown in figure 1 for the cases when the grain of the wood is parallel and perpendicular to the tension. The distribution of $\xi\xi$ on the edge of an isotropic sheet is shown for comparison. The distribution of $\hat{x}\hat{x}$ along the section of the beam between the notches is shown in figure 2.

When the tension is applied parallel to the grain it is seen that the maximum stress, which occurs at the bottom of the notch, rises to $3.5T$, where T is the average tension across the section of the plate between the notches, but that the tensile stress rises only to $1.2T$ when the tension is applied perpendicular to the grain. In the first case, however, the high tensile

TABLE 2. TENSION PROBLEM. VALUES OF STRESSES ON THE EDGE OF THE PLATE: SPRUCE

x/a	y/a	\widehat{xy}/T	\widehat{xx}/T	\widehat{yy}/T	$\widehat{\xi\xi}/T$
Tension parallel to grain					
0	1	0	3.52	0	3.52
0.1001	1.005	0.294	2.95	0.0293	2.98
0.1418	1.01	0.356	2.54	0.0500	2.59
0.201	1.02	0.387	1.96	0.0763	2.04
0.2857	1.04	0.366	1.33	0.101	1.43
0.3516	1.06	0.329	0.992	0.109	1.10
0.4079	1.08	0.295	0.782	0.112	0.894
0.4583	1.1	0.267	0.640	0.111	0.752
0.6633	1.2	0.178	0.323	0.0985	0.421
0.9798	1.4	0.108	0.154	0.0753	0.229
1.249	1.6	0.0780	0.0999	0.0609	0.161
1.497	1.8	0.0618	0.0743	0.0513	0.126
1.732	2.0	0.0514	0.0594	0.0445	0.104
Tension perpendicular to grain					
0	1	0	1.21	0	1.21
0.1001	1.005	0.119	1.19	0.0118	1.20
0.1418	1.01	0.165	1.18	0.0232	1.20
0.201	1.02	0.227	1.15	0.0447	1.20
0.2857	1.04	0.303	1.10	0.0831	1.19
0.3516	1.06	0.360	1.06	0.116	1.17
0.4079	1.08	0.383	1.01	0.145	1.16
0.4583	1.1	0.407	0.977	0.170	1.15
0.6633	1.2	0.456	0.826	0.252	1.08
0.9798	1.4	0.446	0.637	0.312	0.949
1.249	1.6	0.408	0.523	0.319	0.842
1.497	1.8	0.371	0.446	0.308	0.754
1.732	2.0	0.338	0.390	0.292	0.682

TABLE 3. VALUES OF STRESSES AT THE SMALLEST SECTION OF THE BEAM

η°	y/a	\widehat{xx}/T	\widehat{yy}/T	\widehat{xx}/T	\widehat{yy}/T
		tension parallel to grain		tension perpendicular to grain	
45	1	3.52	0	1.21	0
40	0.909	1.64	0.0879	1.14	0.0961
35	0.811	1.20	0.0972	1.08	0.175
30	0.707	0.993	0.0974	1.04	0.239
25	0.598	0.873	0.0956	0.998	0.290
20	0.484	0.797	0.0935	0.968	0.330
15	0.366	0.747	0.0917	0.946	0.359
10	0.246	0.716	0.0904	0.930	0.360
5	0.123	0.698	0.0895	0.921	0.392
0	0	0.693	0.0892	0.918	0.396

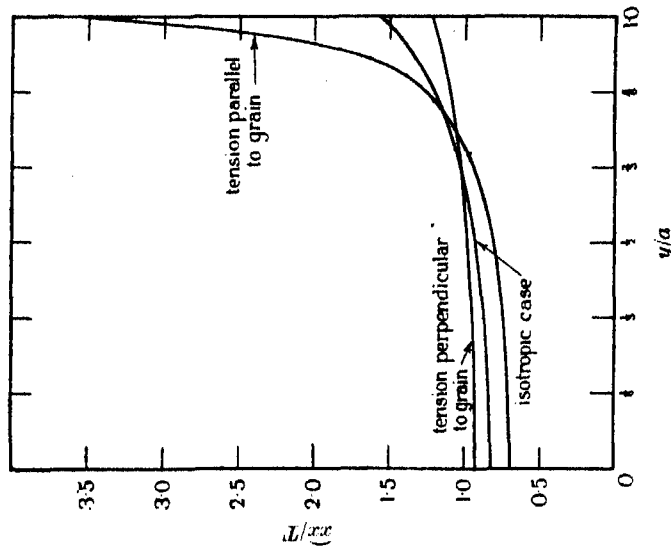


FIGURE 2. Stress distribution across the smallest section of the beam for tension problem.

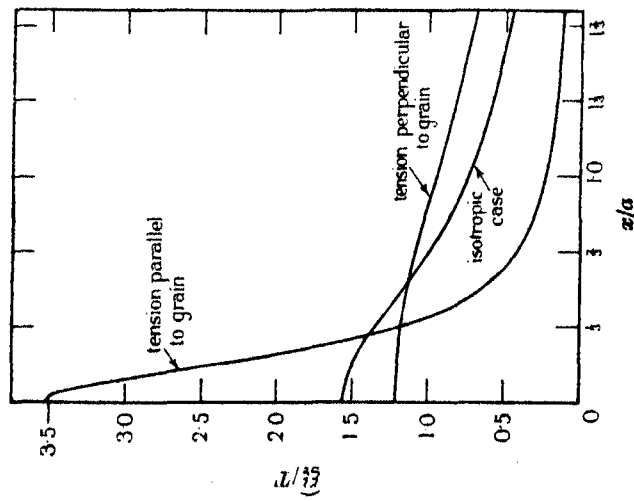


FIGURE 1. Stress distribution along the edge of the plate for tension problem.

stress occurs only in a very small region along the edge of the plate near the bottom of the notches. It also extends only a very short distance into the plate as can be seen from figure 2.

Very little information is at present available about the conditions of failure in stressed wooden members. In a previous paper (Green & Taylor 1945) an attempt was made to estimate the type of failure which might be expected near a hole in a highly stressed spruce plank under tension, and it appears to be of sufficient interest to apply the same process to the present problem in order to estimate the type of failure to be expected near a notch in the edge of a spruce plank under tension. For this purpose measurements of ultimate yield stresses are needed. These are included in a report by C. F. Jenkin (1920), and the relevant figures for spruce are given in table 4. These figures, like those given in table 1 for the elastic constants, are average figures taken from the results of experiments with a large number of specimens which exhibit considerable variation among themselves.

The method consists in comparing the loads at which failure of the five types contemplated in table 4 might be expected to take place in the neighbourhood of a notch in a spruce plank under tensile or compressive loads. For this purpose the value of the externally applied tensile or compressive stress T , which would theoretically raise the stress along the edge of the notch to its ultimate value has been calculated. Four cases are considered:

- (i) tension applied parallel to the grain,
- (ii) compression applied parallel to the grain,
- (iii) tension applied perpendicular to the grain,
- (iv) compression applied perpendicular to the grain,

and for each case the value of T is calculated for each of the five types of failure referred to in table 4, namely,

- (a) rupture of the longitudinal fibres,
- (b) breakdown of the longitudinal fibres in compression,
- (c) rupture perpendicular to the fibres,
- (d) breakdown by compression perpendicular to the fibres,
- (e) shearing rupture parallel to the fibres.

Taking case (i) (a) the stress maximum $\hat{x}\hat{x}$ occurs at the bottom of the notch and is equal to $3.5T$. From table 4 this corresponds with 18,000 lb., so that the average tensile load necessary to cause breakdown of type (a) in case (i) is

$$T = 18,000/3.52 = 5110 \text{ lb./sq.in.}$$

In each case the appropriate maximum positive or negative values of $\hat{x}\hat{x}$, $\hat{y}\hat{y}$, $\hat{x}\hat{y}$ have been taken from table 2. The results are given in table 5.

TABLE 4. ULTIMATE YIELD STRESSES FOR SPRUCE

	lb./sq.in.
tension parallel to the longitudinal fibres	18,000
compression parallel to the fibres	5,000
(corresponding in the present work with \widehat{xx} when the external load is applied along the grain and \widehat{yy} when it is perpendicular to it)	
tension acting in the perpendicular direction, i.e. perpendicular to a vertical plane through the centre of the tree when standing vertically	400-800
compression in this direction	700
(corresponding with \widehat{yy} when the load is applied along the grain and \widehat{xx} when applied perpendicular to it)	
shear stress across a plane perpendicular to the tangential direction	1,100
(corresponding with \widehat{xy})	

TABLE 5. CALCULATED STRESSES FOR FAILURE
NEAR A NOTCH IN A SPRUCE PLANK

failure type lb./sq. in.	case (i)	case (ii)
a	$\frac{18,000}{3.52} = 5110$	—
b	—	$\frac{5000}{3.52} = 1420$
c	$\frac{400-800}{0.112} = 3570-7140$	—
d	—	$\frac{700}{0.112} = 6250$
e	$\frac{1100}{0.39} = 2820$	$\frac{1100}{0.39} = 2820$
	case (iii)	case (iv)
a	$\frac{18,000^*}{0.32} = 56,200$	—
b	—	$\frac{5000}{0.32} = 15,620$
c	$\frac{400-800}{1.2} = 330-660$	—
d	—	$\frac{700}{1.2} = 580$
e	$\frac{1100}{0.46} = 2400$	$\frac{1100}{0.46} = 2400$

* In this case there is a larger value of \widehat{yy} at the mid-point of the section of the plate between the lowest point of the notches, but it makes very little difference to the result.

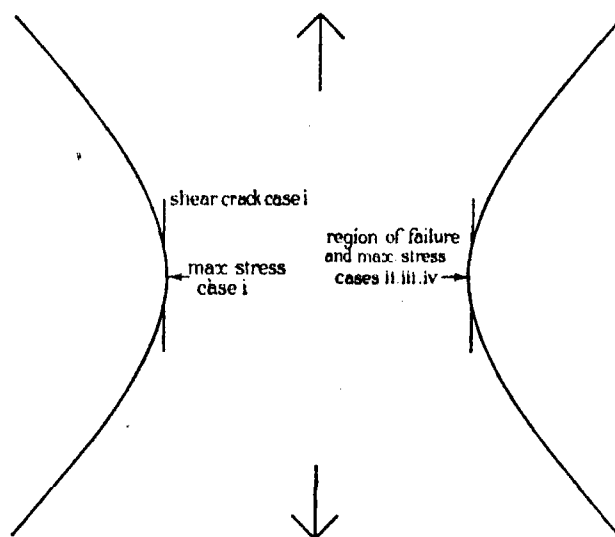


FIGURE 3. Positions of failure for tension or compression.

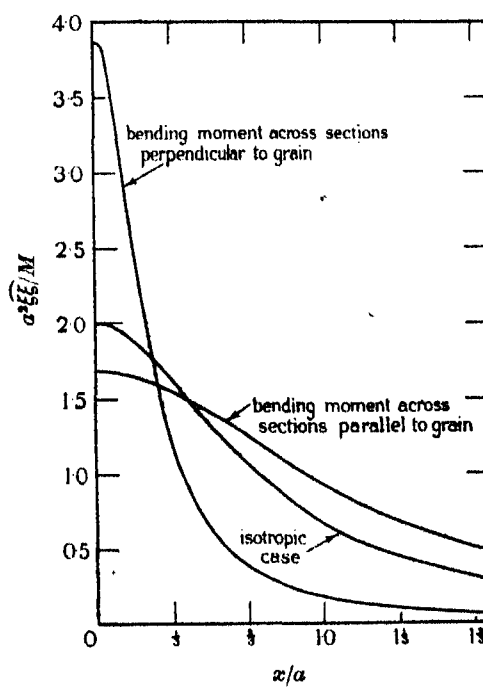


FIGURE 4. Stress distribution along the edge of the plate for bending moment problem.

TABLE 6. FLEXURE PROBLEM. VALUES OF STRESSES ON THE EDGE OF THE PLATE: SPRUCE

x/a	y/a	σ^2_{yy}/M	σ^2_{xx}/M	σ^2_{xy}/M	$\sigma^2_{\xi\xi}/M$
Bending moment across sections perpendicular to grain					
0	1	0	3.87	0	3.87
0.1001	1.005	0.322	3.23	0.0321	3.26
0.1418	1.01	0.387	2.76	0.0544	2.81
0.201	1.02	0.417	2.12	0.0822	2.20
0.2857	1.04	0.387	1.41	0.106	1.51
0.3516	1.06	0.341	1.03	0.113	1.14
0.4079	1.08	0.301	0.796	0.114	0.909
0.4583	1.1	0.267	0.640	0.111	0.751
0.6633	1.2	0.163	0.295	0.0903	0.386
0.9798	1.4	0.0845	0.121	0.0591	0.180
1.249	1.6	0.0536	0.0687	0.0418	0.111
1.497	1.8	0.0377	0.0453	0.0313	0.0767
1.732	2.0	0.0283	0.0326	0.0245	0.0571
Bending moment across sections parallel to grain					
0	1	0	1.69	0	1.69
0.1001	1.005	0.165	1.66	0.0165	1.68
0.1418	1.01	0.229	1.63	0.0322	1.66
0.201	1.02	0.311	1.58	0.0613	1.64
0.2857	1.04	0.407	1.48	0.112	1.59
0.3516	1.06	0.462	1.39	0.153	1.55
0.4079	1.08	0.497	1.31	0.188	1.50
0.4583	1.1	0.518	1.24	0.216	1.46
0.6633	1.2	0.532	0.962	0.294	1.26
0.9798	1.4	0.446	0.637	0.312	0.948
1.249	1.6	0.357	0.457	0.279	0.736
1.497	1.8	0.288	0.347	0.240	0.586
1.732	2.0	0.236	0.273	0.205	0.477

The type of failure which might be expected to occur corresponds with the least of the values given in table 5 for each of the four cases. These are indicated in table 5 by heavy type. It will be noticed that in cases (ii), (iii) and (iv) failure may be expected at the base of the notch, and this is the position of maximum stress. Case (i) is different. Here the type of break-down to be expected is by shearing parallel to the grain. The calculated position of the shear cracks is shown in figure 3.

THE FLEXURE PROBLEM

6. The values of the stresses along the edges of the plate have been evaluated by using formulae (2.4), (2.9), (4.3) and (4.4), and the results are given in table 6. The distribution of $\xi\xi$ over one-quarter of the edge of the

plate is shown in figure 4 for the cases when the grain of the wood is parallel or perpendicular to sections across which there is a bending moment M . The distribution of $\xi\xi$ on the edge of an isotropic sheet is shown for comparison.

When the bending moment acts across sections which are perpendicular to the grain it is seen that the maximum stress, which occurs at the bottom of the notch, rises to $\pm 3.87M/a^2$, but that when the bending moment acts across sections which are parallel to the grain the maximum stress rises only to $\pm 1.69M/a^2$ although the position of the maximum is again at the bottom of the notch. The corresponding maximum for an isotropic material is $\pm 2M/a^2$, and the maximum stress obtained by using the ordinary prismatic beam formula is $\pm 1.5M/a^2$ for both isotropic and aeolotropic materials.

REFERENCES

- Coker, E. G. & Filon, L. N. G. 1931 *A treatise on photo-elasticity*. Camb. Univ. Press.
 Green, A. E. 1939 *Proc. Roy. Soc. A*, **173**, 173.
 Green, A. E. 1942 *Proc. Roy. Soc. A*, **180**, 173.
 Green, A. E. 1945 *Proc. Roy. Soc. A*, **184**, 231.
 Green, A. E. & Taylor, G. I. 1939 *Proc. Roy. Soc. A*, **173**, 162.
 Green, A. E. & Taylor, G. I. 1945 *Proc. Roy. Soc. A*, **184**, 181.
 Jenkin, C. F. 1920 *Report on materials of construction used in aircraft and aircraft engines*. London: Stationery Office.
 Maunsell, F. G. 1936 *Phil. Mag.* **21**, 765.
 Neuber, H. 1933 *Z. angew. Math. Mech.* **13**, 439.
 Neuber, H. 1934 *Ingen.-Arch.* **5**, 238.
 Neuber, H. 1935 *Ingen.-Arch.* **6**, 133.

Stress systems in aeolotropic plates. VII

By A. E. GREEN

(Communicated by G. R. Goldsbrough, F.R.S.—Received 8 September 1943)

The paper is divided into three parts. In Part I a general method of solution is given for problems of stress distributions in isotropic and aeolotropic plates containing a hole of a fairly general shape when the boundary values of the displacements are prescribed, but applications for aeolotropic plates are at present limited to circular and elliptical holes. Part II is concerned with distributions of stress in aeolotropic plates containing a single circular discontinuity or hole when the boundary conditions at the edge of the circle are of a mixed type, i.e. they involve both the displacements and the stresses. A set of fundamental stress functions is obtained which are used to investigate stresses due to bolts and knots in stressed planks, and the work is illustrated numerically for a specimen of spruce wood. In Part III a general method of solution is given for problems of stress distributions in aeolotropic plates which contain any number of circular holes of varying sizes distributed in any manner. The solution is applied to the problem of two circular holes in a tension member, and a few numerical results are given for spruce.

INTRODUCTION

1. The effect of a single unstressed circular hole on certain distributions of stress in a large aeolotropic plate was discussed at some length in a previous paper (Green 1942). With the help of the analysis developed in that paper the distribution of stress in wooden planks was discussed numerically on the assumption that the material could be considered as homogeneous. The effect on stress distributions in aeolotropic plates of single unstressed elliptical holes was considered in another paper (Green 1945). The distribution of stress around holes of a more general type in an isotropic material was also discussed in the same paper.

So far no consideration has been given to problems of stress distributions in aeolotropic plates which contain holes filled with plugs of a different material which occur, for example, with bolted timbers. The conditions at the edge of the plug are difficult to interpret mathematically, but they will involve the displacements as well as the stresses. Apart from bolts in timbers the main discontinuities in wooden materials are knots, and knots may cause considerable variations in the stresses. Since the distribution and the orientation of knots in wood is so variable, calculations of their effect are not likely to be of great practical use, but such calculations may have some theoretical interest. Here again the conditions at the edge of the knot may require the solution of boundary problems involving the displacements.

Some estimate of the effect of rivets in isotropic tension members has been made theoretically and by photo-elastic methods (see Coker & Filon 1931). The effect of discontinuities in isotropic stressed sheets has been studied theoretically by Gurney (1938) and by Donnell (1941), who assumed circular and elliptical discontinuities respectively. All of these problems involved the boundary values of the displacements as well as the stresses. Muschelishvili (1933) has given a general method for solving stress problems in isotropic sheets containing a single hole when the displacements at the edge of the hole are prescribed. In the first part of the present paper a general method of solution is given for problems of stress distributions in isotropic plates containing a hole of a fairly general shape when the boundary values of the displacements only are prescribed. The method differs from that used by Muschelishvili and it is also extended to apply to anisotropic plates, but applications in the latter case are at present limited to circular or elliptical holes. As in previous papers the method of solution requires the use of Airy's stress function and complex variables. Complex variable methods for solving certain stress problems in isotropic materials have been used by Kolosov, Muschelishvili (1933) and others, and recently in an interesting paper by Stevenson (1945), who has kindly allowed me to see his work. Stevenson avoids using Airy's stress function, but his methods for isotropic materials are closely allied to those used by the present writer and there is little to choose between them. It appears, however, that the use of Airy's stress function has some advantages when the work is extended to anisotropic materials.

At present, problems in which the boundary values of the displacements only are prescribed at the edge of the hole appear to be mainly of theoretical rather than practical interest except when the assumed boundary conditions represent the physical facts in a limiting or approximate manner. A better estimate of the physical conditions is often obtained by considering boundary conditions of a mixed type, i.e. those which involve both stresses and displacements, and some progress in this direction can be made when the hole or discontinuity is circular. If the displacements only at the edge of the circle are prescribed the stress distribution can be deduced from the general analysis, but if the hole is filled with another material so that the boundary conditions involve both displacements and stresses, it is more convenient to develop an alternative method of solution. This work is given in Part II. Some numerical applications are given for a specimen of spruce wood.

The effect of a single circular discontinuity or hole in a large anisotropic plate can usually be represented by analytical expressions which, although complicated for many points of the plate, take relatively simple forms at the

edge of the circle so that the necessary numerical work needed to evaluate the stresses at the circle is not too long or complicated for practical purposes. When such discontinuities occur in groups it is to be expected that an analytical description of the stress will be very complex. Stress distributions in isotropic plates containing unstressed circular holes have been studied by several writers, and a general method for solving such problems has been given by the present writer (Green 1940, 1941). The method consisted in finding groups of fundamental stress functions which satisfy the equations of equilibrium and which give single-valued expressions for the corresponding stresses and displacements, and then combining these functions in an infinite series so as to satisfy the boundary conditions at the edges of the holes. The same method is used in Part III of the present paper in order to estimate the effect of groups of circular holes or discontinuities on various stress distributions in aeolotropic plates. It is found that some care has to be taken in selecting the fundamental stress functions in order to secure the convergence of the series involved, but the functions used in Part II are found to be suitable. These functions are fairly complicated in form, and in any given problem considerable numerical work is required before the analytical results can be interpreted. So far, numerical work has only been carried out for the problem of two equal unstressed circular holes in a spruce plank under tension, the holes being such that the line joining their centres is perpendicular to the grain. Spruce was chosen for the numerical work, since it is frequently used in practice and since it is highly aeolotropic. Also calculations have already been made for a spruce plank pierced by one hole. The amount of calculation required for the case when the line of centres of the holes is parallel to the grain is very much more than that required for the case considered here, but calculations for other types of wood which are less aeolotropic would not be so heavy.

PART I

GENERAL EQUATIONS FOR AEOLOTROPIC MATERIALS

2. The plane of the material is taken to be the $z (= x + iy)$ plane, and the plate contains a hole defined by the curve $\eta = 0$ in the transformation

$$z = F(\zeta) = b_0 e^{-i\zeta} + b_n e^{in\zeta}, \quad \frac{dz}{d\zeta} = F'(\zeta) = J e^{i\phi} = a_0 e^{-i\zeta} + a_n e^{in\zeta}, \quad (2.1)$$

where $a_0 = -ib_0$, $a_n = inb_n$ and where ϕ is the angle between the tangent to the curve $\eta = \text{constant}$ through any point and the x -axis. The point $\eta \rightarrow \infty$ corresponds to the point at infinity in the z -plane. The transformation (2.1)

includes the circle and ellipse as special cases. The plate is supposed to be in a state of generalized plane stress in its own plane so that the cartesian components of mean stress may be expressed in terms of Airy's stress function χ by the equations

$$\bar{x}x = \frac{\partial^2 \chi}{\partial y^2}, \quad \bar{y}y = \frac{\partial^2 \chi}{\partial x^2}, \quad \bar{x}y = -\frac{\partial^2 \chi}{\partial x \partial y}. \quad (2.2)$$

Stress components in the orthogonal co-ordinates (ξ, η) are related to the cartesian components by the expressions

$$\left. \begin{aligned} \xi\xi &= \bar{x}x \cos^2 \phi + \bar{x}y \sin 2\phi + \bar{y}y \sin^2 \phi, \\ \eta\eta &= \bar{x}x \sin^2 \phi - \bar{x}y \sin 2\phi + \bar{y}y \cos^2 \phi, \\ \xi\eta &= \frac{1}{2}(\bar{y}y - \bar{x}x) \sin 2\phi + \bar{x}y \cos 2\phi. \end{aligned} \right\} \quad (2.3)$$

When x and y are the directions of symmetry of the material the function χ satisfies the equation

$$\left(\frac{\partial^2}{\partial x^2} + \alpha_1 \frac{\partial^2}{\partial y^2} \right) \left(\frac{\partial^2}{\partial x^2} + \alpha_2 \frac{\partial^2}{\partial y^2} \right) \chi = 0, \quad (2.4)$$

where $\alpha_1 \alpha_2 = s_{11}/s_{22}, \quad \alpha_1 + \alpha_2 = (s_{66} + 2s_{12})/s_{22}, \quad (2.5)$

the elastic constants $s_{11}, s_{22}, s_{12}, s_{66}$ being defined in the usual way (Green & Taylor 1939). In addition, it is convenient to write

$$\gamma_1 = \frac{\alpha_1^{\frac{1}{2}} - 1}{\alpha_1^{\frac{1}{2}} + 1}, \quad \gamma_2 = \frac{\alpha_2^{\frac{1}{2}} - 1}{\alpha_2^{\frac{1}{2}} + 1}, \quad (2.6)$$

and since it has been shown that α_1 and α_2 , if real, are necessarily positive (Green 1943), γ_1 and γ_2 must lie between -1 and 1 .*

The mean cartesian components of displacements are given by

$$u = (s_{12} - s_{11}) \frac{\partial \chi}{\partial x} + s_{11} \frac{\partial \psi}{\partial y}, \quad v = (s_{12} - s_{22}) \frac{\partial \chi}{\partial y} + s_{22} \frac{\partial \psi}{\partial x}, \quad (2.7)$$

where ψ is to be found from the equations

$$\frac{\partial^2 \psi}{\partial x \partial y} = \nabla_1^2 \chi, \quad \frac{\partial^2 \psi}{\partial x^2} + \alpha_1 \alpha_2 \frac{\partial^2 \psi}{\partial y^2} = (1 - \alpha_1)(1 - \alpha_2) \frac{\partial^2 \chi}{\partial x \partial y}. \quad (2.8)$$

* α_1, α_2 may also be complex conjugates, which was not mentioned in Green 1943, and this case can also be covered by similar analysis. In practical problems, however, α_1, α_2 usually appear to be real and therefore positive.

THE GENERAL SOLUTION FOR AEOLOTROPIC MATERIALS

3. A general solution of equation (2.4) may be taken to be the real part of

$$\chi = f(z + \gamma_1 \bar{z}) + g(z + \gamma_2 \bar{z}), \quad (3.1)$$

where $\bar{z} = x - iy$. In previous work terms which produce an isolated force at the origin were separated out from the general functions, but here the function will be left in the form (3.1). From (2.2) and (3.1) it is found that the cartesian components of stress are the real parts of

$$\left. \begin{aligned} \bar{x}\bar{x} &= -(1 - \gamma_1)^2 f''(z + \gamma_1 \bar{z}) - (1 - \gamma_2)^2 g''(z + \gamma_2 \bar{z}), \\ \bar{y}\bar{y} &= (1 + \gamma_1)^2 f''(z + \gamma_1 \bar{z}) + (1 + \gamma_2)^2 g''(z + \gamma_2 \bar{z}), \\ \bar{x}\bar{y} &= -i(1 - \gamma_1^2) f''(z + \gamma_1 \bar{z}) - i(1 - \gamma_2^2) g''(z + \gamma_2 \bar{z}), \end{aligned} \right\} \quad (3.2)$$

and (2.1) and (2.3) then give

$$\left. \begin{aligned} \xi\xi &= -\frac{\{F'(\zeta) - \gamma_1 \bar{F}'(\bar{\zeta})\}^2}{F'(\zeta) \bar{F}'(\bar{\zeta})} f''(z + \gamma_1 \bar{z}) - \frac{\{F'(\zeta) - \gamma_2 \bar{F}'(\bar{\zeta})\}^2}{F'(\zeta) \bar{F}'(\bar{\zeta})} g''(z + \gamma_2 \bar{z}), \\ \eta\eta &= \frac{\{F'(\zeta) + \gamma_1 \bar{F}'(\bar{\zeta})\}^2}{F'(\zeta) \bar{F}'(\bar{\zeta})} f''(z + \gamma_1 \bar{z}) + \frac{\{F'(\zeta) + \gamma_2 \bar{F}'(\bar{\zeta})\}^2}{F'(\zeta) \bar{F}'(\bar{\zeta})} g''(z + \gamma_2 \bar{z}), \\ \xi\eta &= -i \left\{ \frac{F'(\zeta)}{\bar{F}'(\bar{\zeta})} - \frac{\gamma_1^2 \bar{F}'(\bar{\zeta})}{F'(\zeta)} \right\} f''(z + \gamma_1 \bar{z}) - i \left\{ \frac{F'(\zeta)}{\bar{F}'(\bar{\zeta})} - \frac{\gamma_2^2 \bar{F}'(\bar{\zeta})}{F'(\zeta)} \right\} g''(z + \gamma_2 \bar{z}), \end{aligned} \right\} \quad (3.3)$$

where the real parts only are to be taken.

From (2.7), (2.8) and (3.1) the cartesian components of displacement are found to be (see Green 1942) the real parts of

$$\left. \begin{aligned} u &= (1 + \gamma_1)(s_{12} - \alpha_2 s_{22}) f'(z + \gamma_1 \bar{z}) + (1 + \gamma_2)(s_{12} - \alpha_1 s_{22}) g'(z + \gamma_2 \bar{z}), \\ v &= i(1 - \gamma_1)(s_{12} - \alpha_1 s_{22}) f'(z + \gamma_1 \bar{z}) + i(1 - \gamma_2)(s_{12} - \alpha_2 s_{22}) g'(z + \gamma_2 \bar{z}). \end{aligned} \right\} \quad (3.4)$$

THE STRESS RESULTANTS ROUND A CIRCUIT

4. The x and y components of stress at the curve $\eta = \text{constant}$ are

$$\bar{\eta}\bar{x} = -\bar{x}\bar{x} \sin \phi + \bar{x}\bar{y} \cos \phi, \quad \bar{\eta}\bar{y} = \bar{y}\bar{y} \cos \phi - \bar{x}\bar{y} \sin \phi, \quad (4.1)$$

and using (3.2) these become the real parts of

$$\left. \begin{aligned} \bar{\eta}\bar{x} &= -i(1 - \gamma_1)(e^{i\phi} + \gamma_1 e^{-i\phi}) f''(z + \gamma_1 \bar{z}) - i(1 - \gamma_2)(e^{i\phi} + \gamma_2 e^{-i\phi}) g''(z + \gamma_2 \bar{z}), \\ \bar{\eta}\bar{y} &= (1 + \gamma_1)(e^{i\phi} + \gamma_1 e^{-i\phi}) f''(z + \gamma_1 \bar{z}) + (1 + \gamma_2)(e^{i\phi} + \gamma_2 e^{-i\phi}) g''(z + \gamma_2 \bar{z}). \end{aligned} \right\} \quad (4.2)$$

Also, if ds is an element of the curve measured in an anti-clockwise direction

$$\frac{dz}{ds} = -e^{i\phi}, \quad \frac{d\bar{z}}{ds} = -e^{-i\phi}, \quad (4.3)$$

and hence the cartesian components X , Y of the force resultant exerted upon the material outside the curve are the real parts of

$$\begin{aligned} X &= -i(1-\gamma_1) \int f''(z+\gamma_1\bar{z}) d(z+\gamma_1\bar{z}) - i(1-\gamma_2) \int g''(z+\gamma_2\bar{z}) d(z+\gamma_2\bar{z}), \\ Y &= (1+\gamma_1) \int f''(z+\gamma_1\bar{z}) d(z+\gamma_1\bar{z}) + (1+\gamma_2) \int g''(z+\gamma_2\bar{z}) d(z+\gamma_2\bar{z}), \end{aligned} \quad (4.4)$$

or the real parts of the change in values of

$$\begin{aligned} X &= -i(1-\gamma_1)f'(z+\gamma_1\bar{z}) - i(1-\gamma_2)g'(z+\gamma_2\bar{z}), \\ Y &= (1+\gamma_1)f'(z+\gamma_1\bar{z}) + (1+\gamma_2)g'(z+\gamma_2\bar{z}), \end{aligned} \quad (4.5)$$

in passing once round the curve. The couple M exerted upon the material outside the curve $\eta = \text{constant}$ is the real part of

$$M = \int (z+\gamma_1\bar{z})f''(z+\gamma_1\bar{z}) d(z+\gamma_1\bar{z}) + \int (z+\gamma_2\bar{z})g''(z+\gamma_2\bar{z}) d(z+\gamma_2\bar{z}), \quad (4.6)$$

or the real part of the change in value of

$$M = (z+\gamma_1\bar{z})f'(z+\gamma_1\bar{z}) - f(z+\gamma_1\bar{z}) + (z+\gamma_2\bar{z})g'(z+\gamma_2\bar{z}) - g(z+\gamma_2\bar{z}), \quad (4.7)$$

in passing once round the curve.

THE GENERAL SOLUTION FOR ISOTROPIC MATERIALS

5. The stress function is now the real part of

$$\chi = f(z) + \bar{z}g(z), \quad (5.1)$$

and, as in the anisotropic case, terms which produce an isolated force at the origin are included in the general form instead of being displayed separately as in previous work. The corresponding stresses are the real parts of

$$\begin{aligned} \bar{x}\bar{x} &= -f''(z) + 2g'(z) - \bar{F}(\bar{\zeta})g''(z), \\ \bar{y}\bar{y} &= f''(z) + 2g'(z) + \bar{F}(\bar{\zeta})g''(z), \\ \bar{x}\bar{y} &= -if''(z) - i\bar{F}(\bar{\zeta})g''(z), \end{aligned} \quad (5.2)$$

and

$$\begin{aligned} \xi\xi &= 2g'(z) - F'(\zeta)\{f''(z) + \bar{F}(\bar{\zeta})g''(z)\}/\bar{F}'(\bar{\zeta}), \\ \eta\eta &= 2g'(z) + F'(\zeta)\{f''(z) + \bar{F}(\bar{\zeta})g''(z)\}/\bar{F}'(\bar{\zeta}), \\ \xi\eta &= -iF'(\zeta)\{f''(z) + \bar{F}(\bar{\zeta})g''(z)\}/\bar{F}'(\bar{\zeta}). \end{aligned} \quad (5.3)$$

The mean displacements are now the real parts of

$$\left. \begin{aligned} Eu &= (3-\sigma)g(z) - (1+\sigma)\{f'(z) + \bar{z}g'(z)\}, \\ Ev &= -i(3-\sigma)g(z) - i(1+\sigma)\{f'(z) + \bar{z}g'(z)\}, \end{aligned} \right\} \quad (5.4)$$

where E is Young's modulus and σ is the ordinary Poisson's ratio.

The force and couple resultants are found by a similar process to that used in § 4 and are the real parts of

$$\left. \begin{aligned} X &= -i \int \{f''(z) + \bar{z}g''(z)\} dz + i \int g'(z) (dz - d\bar{z}), \\ Y &= \int \{f''(z) + \bar{z}g''(z)\} dz + \int g'(z) (dz + d\bar{z}), \\ M &= \int \{f''(z) + \bar{z}g''(z)\} z dz + \int g'(z) (\bar{z} dz + z d\bar{z}), \end{aligned} \right\} \quad (5.5)$$

or the real parts of the change in values of

$$\left. \begin{aligned} X &= -if'(z) + ig(z) - i\bar{z}g'(z), & Y &= f'(z) + g(z) + \bar{z}g'(z), \\ M &= zf'(z) - f(z) + z\bar{z}g'(z), \end{aligned} \right\} \quad (5.6)$$

in passing once round the curve. It can immediately be verified that (4.5) and (5.6) give the results (4.17), (2.23) and (2.24) of Paper IV (Green 1942).

SOLUTION FOR GIVEN EDGE DISPLACEMENTS

6. When the displacements have prescribed values at the edge of the hole $\eta = \text{constant}$ it is convenient to introduce two functions of ζ , $G(\zeta)$ and $H(\zeta)$, which are such that the real part of $G(\zeta) = u_e$ and the imaginary part of $H(\zeta) = v_e$ on the edge of the hole. u_e and v_e represent the edge values of the cartesian components of the displacement. From (3.4) it is seen that $G(\zeta)$ and $H(\zeta)$ may be written in the form

$$\left. \begin{aligned} G(\zeta) &= (1+\gamma_1)(s_{12}-\alpha_2s_{22})f'\{F(\zeta)+\gamma_1\bar{F}(\zeta)\} \\ &\quad + (1+\gamma_2)(s_{12}-\alpha_1s_{22})g'\{F(\zeta)+\gamma_2\bar{F}(\zeta)\}, \\ H(\zeta) &= -(1-\gamma_1)(s_{12}-\alpha_1s_{22})f'\{F(\zeta)+\gamma_1\bar{F}(\zeta)\} \\ &\quad - (1-\gamma_2)(s_{12}-\alpha_2s_{22})g'\{F(\zeta)+\gamma_2\bar{F}(\zeta)\}, \end{aligned} \right\} \quad (6.1)$$

for aeolotropic materials. Using (5.4) the corresponding forms for isotropic materials are found to be

$$\left. \begin{aligned} EG(\zeta) &= (3-\sigma)g(z) - (1+\sigma)\{f'(z) + \bar{F}(\zeta)g'(z)\}, \\ EH(\zeta) &= (3-\sigma)g(z) + (1+\sigma)\{f'(z) + \bar{F}(\zeta)g'(z)\}. \end{aligned} \right\} \quad (6.2)$$

For isotropic materials it is possible to solve equations (6.2) for the functions $f(z)$, $g(z)$ and then to obtain general expressions for the stresses from (5.2) and (5.3). When the hole in an aeolotropic material is circular or elliptical general values for the stresses may be found from (6.1), (3.2) and (3.3). All these stresses are fairly complicated but they take comparatively simple forms at the edge of the hole and only these values will be given here. Thus the edge stresses for an aeolotropic material are found to be the real parts of

$$2\mu F'(\zeta) \bar{F}'(\zeta) \widehat{\eta\eta_e} = G'(\zeta) \{\rho_1 F'(\zeta) + \rho_2 \bar{F}'(\zeta)\} - H'(\zeta) \{\rho_3 F'(\zeta) - \rho_4 \bar{F}'(\zeta)\}, \quad (6.3)$$

$$2\mu F'(\zeta) \bar{F}'(\zeta) \widehat{\xi\xi_e} = -iG'(\zeta) \{\rho_1 F'(\zeta) - \rho_2 \bar{F}'(\zeta)\} + iH'(\zeta) \{\rho_3 F'(\zeta) + \rho_4 \bar{F}'(\zeta)\}, \quad (6.4)$$

$$\begin{aligned} \mu(\widehat{\xi\xi_e} + \widehat{\eta\eta_e}) &= 2[G'(\zeta) \{\rho_2 F'(\zeta) + \gamma_1 \gamma_2 \rho_1 \bar{F}'(\zeta)\} + H'(\zeta) \{\rho_4 F'(\zeta) - \gamma_1 \gamma_2 \rho_3 \bar{F}'(\zeta)\}] \\ &\quad \times \{F'(\zeta) + \gamma_1 \bar{F}'(\zeta)\}^{-1} \{F'(\zeta) + \gamma_2 \bar{F}'(\zeta)\}^{-1}, \end{aligned} \quad (6.5)$$

where

$$\left. \begin{aligned} \rho_1 &= s_{12} + \frac{s_{22}(3 + \gamma_1 + \gamma_2 - \gamma_1 \gamma_2)}{(1 - \gamma_1)(1 - \gamma_2)}, & \rho_2 &= s_{12} - \frac{s_{22}(1 - \gamma_1 - \gamma_2 - 3\gamma_1 \gamma_2)}{(1 - \gamma_1)(1 - \gamma_2)}, \\ \rho_3 &= s_{12} + \frac{s_{11}(3 - \gamma_1 - \gamma_2 - \gamma_1 \gamma_2)}{(1 + \gamma_1)(1 + \gamma_2)}, & \rho_4 &= s_{12} - \frac{s_{11}(1 + \gamma_1 + \gamma_2 - 3\gamma_1 \gamma_2)}{(1 + \gamma_1)(1 + \gamma_2)}, \\ \mu &= s_{12}^2 + \frac{2s_{12}s_{22}(1 + \gamma_1)(1 + \gamma_2)}{(1 - \gamma_1)(1 - \gamma_2)} - \frac{s_{22}^2(1 + \gamma_1)(1 + \gamma_2)}{(1 - \gamma_1)^3(1 - \gamma_2)^3} \{3(1 - \gamma_1 \gamma_2)^2 + (\gamma_1 - \gamma_2)^2\}, \\ &= \frac{1}{2}(\rho_1 \rho_4 + \rho_2 \rho_3). \end{aligned} \right\} \quad (6.6)$$

For isotropic materials the edge stresses are the real parts of (6.3) and (6.4) where the constants now have the values

$$\rho_1 = \rho_3 = (3 - \sigma)/E, \quad \rho_2 = \rho_4 = -(1 + \sigma)/E, \quad \mu = \rho_1 \rho_2 = \rho_3 \rho_4. \quad (6.7)$$

The expression (6.5) simplifies considerably for the isotropic case and becomes the real part of

$$(3 - \sigma)(\widehat{\xi\xi_e} + \widehat{\eta\eta_e}) = 2E\{G'(\zeta) + H'(\zeta)\}/F'(\zeta). \quad (6.8)$$

7. Only one application of the above analysis will be considered here. It will be supposed that the origin of co-ordinates is inside the hole and that each point of the boundary of the hole receives a small rigid body rotation about the origin through an angle β . Stevenson and others have solved this problem for the case of an elliptical hole in an isotropic material. The

present work is applicable to circular or elliptical holes in aeolotropic materials or for the more general hole (2.1) in isotropic materials. The values of the displacement components at the edge of the hole are

$$u_e = -\beta y, \quad v_e = \beta x, \quad (7.1)$$

and the corresponding values of $G(\zeta)$, $H(\zeta)$ are

$$\begin{aligned} G(\zeta) &= i\beta\{F(\zeta) - b_0 e^{-i\zeta} - \bar{b}_0 e^{i\zeta}\}, \\ H(\zeta) &= i\beta\{F(\zeta) - b_0 e^{-i\zeta} + \bar{b}_0 e^{i\zeta}\}. \end{aligned} \quad (7.2)$$

When the hole is given by the transformation (2.1) it can easily be verified that (7.2) are suitable forms for $G(\zeta)$ and $H(\zeta)$ for isotropic materials, giving vanishing stresses and displacements at infinity, whilst from (5.6) and (6.2) it is found that the couple applied to the material outside the hole is

$$M = 2\pi\beta E \left\{ \frac{b_0 \bar{b}_0}{1+\sigma} + \frac{nb_n \bar{b}_n}{3-\sigma} \right\}. \quad (7.3)$$

For aeolotropic materials the forms (7.2) are suitable when the hole is circular or elliptical, that is when n is 0 or 1 in (2.1). The couple is now obtained from (4.7), (6.1) and (7.2) and is found to be the real part of

$$M = \pi\beta\{(b_1 - \bar{b}_0)(\rho_1 b_0 + \rho_2 \bar{b}_1) - (b_1 + \bar{b}_0)(\rho_3 b_0 - \rho_4 \bar{b}_1)\}/\mu. \quad (7.4)$$

PART II

GENERAL STRESS FUNCTIONS FOR CIRCULAR HOLES

8. When the problem of an aeolotropic tension member containing a circular hole was first solved (Green & Taylor 1945) fundamental stress functions were used which were equivalent to the complex stress functions

$$(z + \gamma_1 \bar{z})^{-n}, \quad (z + \gamma_2 \bar{z})^{-n} \quad (n \geq 1), \quad (8.1)$$

these being the most obvious extension of the forms used for isotropic materials. It was found, however, that these were not completely suitable as, when they were combined in an infinite series, the series did not converge at every point of the plate. A general method for finding suitable stress functions was given in a later paper (Green 1942). Two functions of z , $V(z)$ and $W(z)$, were defined which were such that the real part of $V(z) = -\hat{r}\hat{r}_e$ and the imaginary part of $W(z) = -\hat{r}\hat{\theta}_e$ at the edge of the hole $r = a$. The

general stress functions were then obtained in terms of $V(z)$ and $W(z)$ by means of the equations

$$\begin{aligned}
 (\gamma_1 - \gamma_2) \left(1 - \frac{4\gamma_1 a^2}{u^2}\right)^{\frac{1}{2}} \{uf''(u) + A(1 - \gamma_2)(s_{12} - \alpha_2 s_{22}) + iB(1 + \gamma_2)(s_{12} - \alpha_1 s_{22})\} \\
 = A(1 - \gamma_2)^2(s_{12} - \alpha_2 s_{22}) - A(1 - \gamma_1)(1 - \gamma_2)(s_{12} - \alpha_1 s_{22}) \\
 - iB(1 + \gamma_2)^2(s_{12} - \alpha_1 s_{22}) + iB(1 + \gamma_1)(1 + \gamma_2)(s_{12} - \alpha_2 s_{22}) \\
 - \frac{1}{4}u\{V(u) + W(u)\} \left\{1 + \left(1 - \frac{4\gamma_1 a^2}{u^2}\right)^{\frac{1}{2}}\right\} \\
 - \frac{\gamma_2 u}{4\gamma_1} \{V(u) - W(u)\} \left\{1 - \left(1 - \frac{4\gamma_1 a^2}{u^2}\right)^{\frac{1}{2}}\right\}, \quad (8.2)
 \end{aligned}$$

$$\begin{aligned}
 (\gamma_1 - \gamma_2) \left(1 - \frac{4\gamma_2 a^2}{v^2}\right)^{\frac{1}{2}} \{vg''(v) - A(1 - \gamma_1)(s_{12} - \alpha_1 s_{22}) - iB(1 + \gamma_1)(s_{12} - \alpha_2 s_{22})\} \\
 = A(1 - \gamma_1)^2(s_{12} - \alpha_1 s_{22}) - A(1 - \gamma_1)(1 - \gamma_2)(s_{12} - \alpha_2 s_{22}) \\
 - iB(1 + \gamma_1)^2(s_{12} - \alpha_2 s_{22}) + iB(1 + \gamma_1)(1 + \gamma_2)(s_{12} - \alpha_1 s_{22}) \\
 + \frac{1}{4}v\{V(v) + W(v)\} \left\{1 + \left(1 - \frac{4\gamma_2 a^2}{v^2}\right)^{\frac{1}{2}}\right\} \\
 + \frac{\gamma_1 v}{4\gamma_2} \{V(v) - W(v)\} \left\{1 - \left(1 - \frac{4\gamma_2 a^2}{v^2}\right)^{\frac{1}{2}}\right\}, \quad (8.3)
 \end{aligned}$$

where $V(u)$, $V(v)$, $W(u)$, $W(v)$ are found from $V(z)$, $W(z)$ by using the substitutions

$$z = \frac{1}{2}u \left\{1 + \left(1 - \frac{4\gamma_1 a^2}{u^2}\right)^{\frac{1}{2}}\right\}, \quad z = \frac{1}{2}v \left\{1 + \left(1 - \frac{4\gamma_2 a^2}{v^2}\right)^{\frac{1}{2}}\right\}, \quad (8.4)$$

and afterwards u and v are given by

$$u = z + \gamma_1 \bar{z}, \quad v = z + \gamma_2 \bar{z}. \quad (8.5)$$

The constants A , B are related to the cartesian components X , Y of force at the origin by the equations

$$\begin{aligned}
 X &= 2\pi A(1 - \gamma_1)(1 - \gamma_2)(\alpha_1 - \alpha_2)s_{22}, \\
 Y &= 2\pi B(1 + \gamma_1)(1 + \gamma_2)(\alpha_1 - \alpha_2)s_{22}.
 \end{aligned} \quad (8.6)$$

The stresses corresponding to (8.2) and (8.3) are found from (3.2) and (3.3) by using the transformation

$$z = ae^{-i\theta}. \quad (8.7)$$

By choosing suitable values for $V(z)$, $W(z)$ it is possible to find stress functions which have prescribed values for the normal and shear stresses at the edge of the hole $r = a$. A list of functions suitable for the problems in

view is given below, together with the corresponding values of the stresses at the edge of the hole, in plane polar co-ordinates r, θ , and also the corresponding edge values of the cartesian components of the displacement. For convenience the following notation is introduced, namely,

$$X_1 = R_1^\dagger e^{i\phi_1}, \quad X_2 = R_2^\dagger e^{i\phi_2}, \quad (8.8)$$

$$\text{where} \quad R_1 e^{i\phi_1} = 1 - \frac{4a^2\gamma_1}{(z + \gamma_1\bar{z})^2}, \quad R_2 e^{i\phi_2} = 1 - \frac{4a^2\gamma_2}{(z + \gamma_2\bar{z})^2}. \quad (8.9)$$

The quantities R_1 and R_2 are chosen to be positive and ϕ_1 and ϕ_2 are to lie between $-\pi$ and π :

$$\begin{aligned} \chi_n = f_n(u) + g_n(v), \quad (n \geq 0). \quad V(z) = -W(z) = -a^n/z^n, \quad A = B = 0 \\ \left. \begin{aligned} (\gamma_1 - \gamma_2) a^n (2\gamma_1)^{n+1} X_1 f_n''(u) &= \gamma_2 u^n (1 - X_1)^{n+1}, \\ (\gamma_1 - \gamma_2) a^n (2\gamma_2)^{n+1} X_2 g_n''(v) &= -\gamma_1 v^n (1 - X_2)^{n+1}, \end{aligned} \right\} \quad (8.10) \end{aligned}$$

$$u_e = \frac{a\rho_4 \cos(n+1)\theta}{n+1}, \quad v_e = \frac{a\rho_2 \sin(n+1)\theta}{n+1}, \quad (8.11)$$

$$\left. \begin{aligned} \bar{r}\bar{r}_e &= \cos n\theta, \quad \bar{r}\bar{\theta}_e = \sin n\theta, \\ \bar{\theta}\bar{\theta}_e (1 + \gamma_1^2 - 2\gamma_1 \cos 2\theta) (1 + \gamma_2^2 - 2\gamma_2 \cos 2\theta) \\ &= -\gamma_1 \gamma_2 \cos(n-4)\theta + (\gamma_1 + \gamma_2) (1 + \gamma_1 \gamma_2) \cos(n-2)\theta \\ &\quad + \{3\gamma_1^2 \gamma_2^2 - (\gamma_1 + \gamma_2)^2 - 1\} \cos n\theta \\ &\quad + (\gamma_1 + \gamma_2) (1 - 3\gamma_1 \gamma_2) \cos(n+2)\theta + 3\gamma_1 \gamma_2 \cos(n+4)\theta. \end{aligned} \right\} \quad (8.12)$$

$$\begin{aligned} \chi'_n = F_n(u) + G_n(v), \quad (n \geq 2). \quad V(z) = W(z) = -a^n/z^n, \quad A = B = 0 \\ \left. \begin{aligned} (\gamma_1 - \gamma_2) a^{n-2} (2\gamma_1)^{n-1} X_1 F_n''(u) &= u^{n-2} (1 - X_1)^{n-1}, \\ (\gamma_1 - \gamma_2) a^{n-2} (2\gamma_2)^{n-1} X_2 G_n''(v) &= -v^{n-2} (1 - X_2)^{n-1}, \end{aligned} \right\} \quad (8.13) \end{aligned}$$

$$u_e = -\frac{a\rho_3 \cos(n-1)\theta}{n-1}, \quad v_e = \frac{a\rho_1 \sin(n-1)\theta}{n-1}, \quad (8.14)$$

$$\left. \begin{aligned} \bar{r}\bar{r}_e &= \cos n\theta, \quad \bar{r}\bar{\theta}_e = -\sin n\theta, \\ \bar{\theta}\bar{\theta}_e (1 + \gamma_1^2 - 2\gamma_1 \cos 2\theta) (1 + \gamma_2^2 - 2\gamma_2 \cos 2\theta) \\ &= 3\gamma_1 \gamma_2 \cos(n-4)\theta + (\gamma_1 + \gamma_2) (\gamma_1 \gamma_2 - 3) \cos(n-2)\theta \\ &\quad + \{3 - (\gamma_1 + \gamma_2)^2 - \gamma_1^2 \gamma_2^2\} \cos n\theta \\ &\quad + (\gamma_1 + \gamma_2) (1 + \gamma_1 \gamma_2) \cos(n+2)\theta - \gamma_1 \gamma_2 \cos(n+4)\theta. \end{aligned} \right\} \quad (8.15)$$

$$\begin{aligned} \Phi_n = h_n(u) + k_n(v), \quad (n \geq 0). \quad V(z) = -W(z) = -ia^n/z^n, \quad A = B = 0 \\ \left. \begin{aligned} (\gamma_1 - \gamma_2) a^n (2\gamma_1)^{n+1} X_1 h_n''(u) &= i\gamma_2 u^n (1 - X_1)^{n+1}, \\ (\gamma_1 - \gamma_2) a^n (2\gamma_2)^{n+1} X_2 k_n''(v) &= -i\gamma_1 v^n (1 - X_2)^{n+1}, \end{aligned} \right\} \quad (8.16) \end{aligned}$$

$$u_e = \frac{a\rho_4 \sin(n+1)\theta}{n+1}, \quad v_e = -\frac{a\rho_3 \cos(n+1)\theta}{n+1}, \quad (8.17)$$

$$\left. \begin{aligned} \widehat{r}\theta_e &= \sin n\theta, \quad \widehat{r}\partial_e = -\cos n\theta, \\ \partial\theta_e(1+\gamma_1^2-2\gamma_1\cos 2\theta)(1+\gamma_2^2-2\gamma_2\cos 2\theta) \\ &= -\gamma_1\gamma_2\sin(n-4)\theta + (\gamma_1+\gamma_2)(1+\gamma_1\gamma_2)\sin(n-2)\theta \\ &\quad + \{3\gamma_1^2\gamma_2^2 - (\gamma_1+\gamma_2)^2 - 1\}\sin n\theta \\ &\quad + (\gamma_1+\gamma_2)(1-3\gamma_1\gamma_2)\sin(n+2)\theta + 3\gamma_1\gamma_2\sin(n+4)\theta. \end{aligned} \right\} \quad (8.18)$$

$$\Phi'_n = H_n(u) + K_n(v), \quad (n \geq 2). \quad V(z) = W(z) = -ia^n/z^n, \quad A = B = 0$$

$$\left. \begin{aligned} (\gamma_1 - \gamma_2)a^{n-2}(2\gamma_1)^{n-1}X_1H''_n(u) &= iu^{n-2}(1-X_1)^{n-1}, \\ (\gamma_1 - \gamma_2)a^{n-2}(2\gamma_2)^{n-1}X_2K''_n(v) &= -iv^{n-2}(1-X_2)^{n-1}, \end{aligned} \right\} \quad (8.19)$$

$$u_e = -\frac{a\rho_3 \sin(n-1)\theta}{n-1}, \quad v_e = -\frac{a\rho_1 \cos(n-1)\theta}{n-1}, \quad (8.20)$$

$$\left. \begin{aligned} \widehat{r}\theta_e &= \sin n\theta, \quad \widehat{r}\partial_e = \cos n\theta, \\ \partial\theta_e(1+\gamma_1^2-2\gamma_1\cos 2\theta)(1+\gamma_2^2-2\gamma_2\cos 2\theta) \\ &= 3\gamma_1\gamma_2\sin(n-4)\theta + (\gamma_1+\gamma_2)(\gamma_1\gamma_2-3)\sin(n-2)\theta \\ &\quad + \{3 - (\gamma_1+\gamma_2)^2 - \gamma_1^2\gamma_2^2\}\sin n\theta \\ &\quad + (\gamma_1+\gamma_2)(1+\gamma_1\gamma_2)\sin(n+2)\theta - \gamma_1\gamma_2\sin(n+4)\theta. \end{aligned} \right\} \quad (8.21)$$

The above list of fundamental stress functions contains those which produce zero force resultants at the hole. To complete the list two more functions are added which give force resultants of amount $-2\pi a$ along the x - and y -axes respectively:

$$\chi'_1 = F_1(u) + A(1-\gamma_2)(s_{12}-\alpha_2s_{22})u \log u + G_1(v) - A(1-\gamma_1)(s_{12}-\alpha_1s_{22})v \log v$$

$$V(z) = W(z) = -a/z, \quad B = 0, \quad A(1-\gamma_1)(1-\gamma_2)(\alpha_1-\alpha_2)s_{22} = -a, \quad (8.22)$$

$$\left. \begin{aligned} (\alpha_1-\alpha_2)X_1\{uF''_1(u) + A(1-\gamma_2)(s_{12}-\alpha_2s_{22})\} &= -\frac{a(s_{12}-\alpha_2s_{22})}{(1-\gamma_1)s_{22}}, \\ (\alpha_1-\alpha_2)X_2\{vG''_1(v) - A(1-\gamma_1)(s_{12}-\alpha_1s_{22})\} &= \frac{a(s_{12}-\alpha_1s_{22})}{(1-\gamma_2)s_{22}}, \end{aligned} \right\} \quad (8.23)$$

$$\left. \begin{aligned} \widehat{r}\theta_e &= \cos \theta, \quad \widehat{r}\partial_e = -\sin \theta, \\ \partial\theta_e &= -\cos \theta - \frac{4 \cos \theta}{(\alpha_1-\alpha_2)s_{22}} \left\{ \frac{\gamma_1(s_{12}-\alpha_2s_{22})}{1+\gamma_1^2-2\gamma_1\cos 2\theta} - \frac{\gamma_2(s_{12}-\alpha_1s_{22})}{1+\gamma_2^2-2\gamma_2\cos 2\theta} \right\}. \end{aligned} \right\} \quad (8.24)$$

u_e and v_e are constants.

$$\Phi'_1 = H_1(u) + iB(1 + \gamma_2)(s_{12} - \alpha_1 s_{22})u \log u + K_1(v) - iB(1 + \gamma_1)(s_{12} - \alpha_2 s_{22})v \log v$$

$$V(z) = W(z) = -ia/z, \quad A = 0, \quad B(1 + \gamma_1)(1 + \gamma_2)(\alpha_1 - \alpha_2)s_{22} = -a, \quad (8.25)$$

$$\left. \begin{aligned} (\alpha_1 - \alpha_2) X_1 \{u H_1''(u) + iB(1 + \gamma_2)(s_{12} - \alpha_1 s_{22})\} &= -\frac{ia(s_{12} - \alpha_1 s_{22})}{(1 + \gamma_1)s_{22}}, \\ (\alpha_1 - \alpha_2) X_2 \{v K_1''(v) - iB(1 + \gamma_1)(s_{12} - \alpha_2 s_{22})\} &= \frac{ia(s_{12} - \alpha_2 s_{22})}{(1 + \gamma_2)s_{22}}, \end{aligned} \right\} \quad (8.26)$$

$$\left. \begin{aligned} \widehat{r}_e &= \sin \theta, \quad \widehat{\theta}_e = \cos \theta, \\ \widehat{\theta}_e &= -\sin \theta - \frac{4 \sin \theta}{(\alpha_1 - \alpha_2)s_{22}} \left\{ \frac{\gamma_1(s_{12} - \alpha_1 s_{22})}{1 + \gamma_1^2 - 2\gamma_1 \cos 2\theta} - \frac{\gamma_2(s_{12} - \alpha_2 s_{22})}{1 + \gamma_2^2 - 2\gamma_2 \cos 2\theta} \right\}. \end{aligned} \right\} \quad (8.27)$$

u_e and v_e are constants.

9. When the region $r \leq a$ contains a different material from the rest of the plate further stress functions are needed. If the material inside the circle $r = a$ is isotropic, general stress functions may easily be obtained, but if it is aeolotropic, general functions appear to be very complicated. The functions required for problems discussed in this paper are, however, fairly simple and only these will be given here. The material in the region $r = a$ is supposed to have the same directions of symmetry as the material outside the circle and dashes will be added to denote the elastic constants for the material:

$$\Psi_0 = l_0(u) + m_0(v)$$

$$l_0''(u) = \frac{1 + \gamma_2'^2}{2(\gamma_1' - \gamma_2')(1 - \gamma_1'\gamma_2')}, \quad m_0''(v) = -\frac{1 + \gamma_1'^2}{2(\gamma_1' - \gamma_2')(1 - \gamma_1'\gamma_2')}, \quad (9.1)$$

$$\widehat{r} = 1, \quad \widehat{\theta} = 0, \quad \widehat{\theta\theta} = 1, \quad (9.2)$$

$$u = (s'_{11} + s'_{12})r \cos \theta, \quad v = (s'_{22} + s'_{12})r \sin \theta. \quad (9.3)$$

$$\Psi_2 = l_2(u) + m_2(v)$$

$$l_2''(u) = \frac{\gamma_2'}{(\gamma_1' - \gamma_2')(1 - \gamma_1'\gamma_2')}, \quad m_2''(v) = -\frac{\gamma_1'}{(\gamma_1' - \gamma_2')(1 - \gamma_1'\gamma_2')}, \quad (9.4)$$

$$\widehat{r} = \cos 2\theta, \quad \widehat{\theta} = -\sin 2\theta, \quad \widehat{\theta\theta} = -\cos 2\theta, \quad (9.5)$$

$$u = (s'_{11} - s'_{12})r \cos \theta, \quad v = (s'_{12} - s'_{22})r \sin \theta. \quad (9.6)$$

10. In the rest of this part of the paper a number of problems will be examined with the help of the above stress functions. The work will be illustrated numerically for a specimen of spruce wood whose elastic constants are shown in table 1. The constants are chosen so that the grain of the wood is parallel to the y -axis. If the values of s_{11} and s_{22} are interchanged (which means a change in sign of γ_1 and γ_2), then the grain will be

parallel to the x -axis. The wood is cut from the tree so that the annular layers are parallel to the plane of the plank and the elastic constants s_{11} , s_{22} , s_{12} , s_{66} are measured in sq. mm./1000 kg.

TABLE 1

	s_{11}	s_{22}	s_{12}	s_{66}	α_1	α_2	γ_1	γ_2
spruce	15.5	0.587	-0.33	11.5	16.91	1.56	0.608	0.111

NORMAL DISPLACEMENT AT THE EDGE OF THE HOLE

11. In isotropic materials an all-round pressure at the edge of a circular hole produces a uniform normal displacement, but this is not true for aeolotropic materials. The problem of a uniform all-round normal pressure at the edge of a circular hole in an aeolotropic material was solved in a previous paper (Green 1942). A solution is now given for the problem of a uniform normal displacement at the edge of a circular hole. Such a condition arises approximately if a pin or rivet is inserted into a hole, the natural diameter of the pin being greater than the natural diameter of the hole, and when the pin is assumed to be very rigid compared with the material of the plate. Since the problem in this form is one in which the edge displacements have prescribed values, the solution could be deduced from Part I. The solution given here uses the fundamental stress functions developed above. The required stress function is

$$\chi = A\chi_0 + B\chi'_2, \quad (11.1)$$

where the constants A , B are to be found from the condition that each point of the circle $r = a$ receives a small normal displacement c so that

$$a\rho_4 A - a\rho_3 B = c, \quad a\rho_2 A + a\rho_1 B = c. \quad (11.2)$$

The average normal pressure P on the edge of the hole is $-A$ and the edge stresses are

$$\left. \begin{aligned} r\bar{r}_e &= A + B \cos 2\theta, \quad r\bar{\theta}_e = -B \sin 2\theta, \\ \bar{\theta}_e(1 + \gamma_1^2 - 2\gamma_1 \cos 2\theta)(1 + \gamma_2^2 - 2\gamma_2 \cos 2\theta) \\ &= A\{3\gamma_1^2\gamma_2^2 - (\gamma_1 + \gamma_2)^2 - 1 + 2(\gamma_1 + \gamma_2)(1 - \gamma_1\gamma_2) \cos 2\theta + 2\gamma_1\gamma_2 \cos 4\theta\} \\ &\quad + B\{(\gamma_1 + \gamma_2)(\gamma_1\gamma_2 - 3) + [3 + 3\gamma_1\gamma_2 - (\gamma_1 + \gamma_2)^2 - \gamma_1^2\gamma_2^2] \cos 2\theta \\ &\quad + (\gamma_1 + \gamma_2)(1 + \gamma_1\gamma_2) \cos 4\theta - \gamma_1\gamma_2 \cos 6\theta\}. \end{aligned} \right\} \quad (11.3)$$

Using the values of the elastic constants which are given in table 1 it is found that $B = 0.527P$ and the values of the edge stresses are recorded in

table 2 where the line $\theta = 0$ is parallel to the grain. The values of $\widehat{\theta\theta}$, \widehat{rr} are shown in figure 1 and the values of $\widehat{\theta\theta}$ corresponding to constant normal pressure are included for comparison. For an isotropic material the value of $\widehat{\theta\theta}/P$ equals 1 at all points of the circle for both the case of constant normal pressure and the case of constant normal displacement, but there is a marked difference in the two cases when the material is as highly aeolotropic as spruce. The maximum of $\widehat{\theta\theta}$ occurs at very different points in the two cases and is much greater for the case of constant normal displacement.

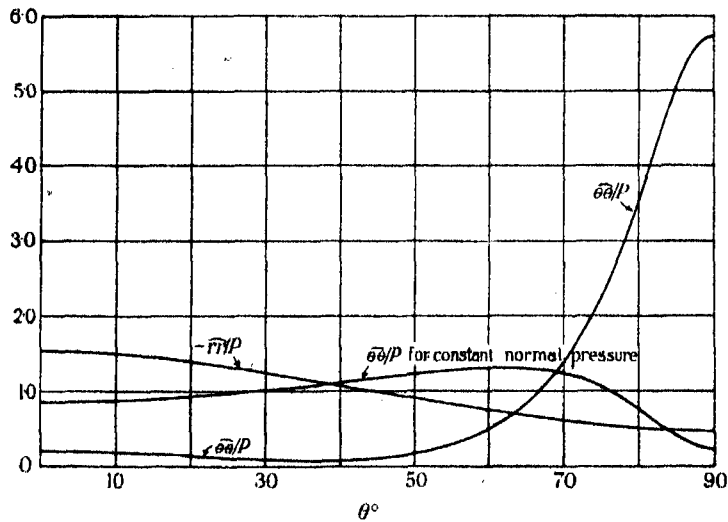


FIGURE 1. Stress distribution at the edge of the circle $r=a$ in spruce.
Grain parallel to $\theta=0$.

TABLE 2. VALUES OF STRESSES AT THE EDGE OF THE CIRCLE: SPRUCE

θ°	\widehat{rr}/P	$\widehat{\theta\theta}/P$	$\widehat{r\theta}/P$	\widehat{xx}/P	\widehat{yy}/P	\widehat{xy}/P
0	-1.53	0.195	0	-1.53	0.195	0
10	-1.50	0.178	-0.180	-1.51	0.190	-0.117
20	-1.40	0.134	-0.339	-1.44	0.172	-0.235
30	-1.26	0.083	-0.457	-1.32	0.142	-0.355
40	-1.09	0.064	-0.519	-1.13	0.098	-0.479
50	-0.908	0.149	-0.519	-0.800	0.040	-0.611
60	-0.738	0.477	-0.457	-0.222	-0.037	-0.754
70	-0.596	1.39	-0.339	0.936	-0.146	-0.897
75	-0.543	2.27	-0.264	1.95	-0.223	-0.933
80	-0.504	3.55	-0.180	3.37	-0.320	-0.864
85	-0.481	5.02	-0.092	4.97	-0.423	-0.568
90	-0.473	5.77	0	5.77	-0.473	0

UNIFORM TANGENTIAL DISPLACEMENT AT THE HOLE

12. This is a special case of the problem solved in Part I. If each point of the circumference of the hole receives a small rotation about the centre through an angle β then a suitable stress function is

$$\chi = A\Phi_0 + B\Phi'_2, \quad (12.1)$$

where $\rho_4 A - \rho_3 B = -\beta, \quad \rho_2 A + \rho_1 B = -\beta. \quad (12.2)$

The couple G required to produce this displacement is $-2\pi a^2 A$ and the corresponding edge stresses are

$$\left. \begin{aligned} \hat{r}r_e &= B \sin 2\theta, \quad \hat{r}\hat{\theta}_e = -A + B \cos 2\theta, \\ \hat{\theta}\hat{\theta}_e &= (1 + \gamma_1^2 - 2\gamma_1 \cos 2\theta)(1 + \gamma_2^2 - 2\gamma_2 \cos 2\theta) \\ &= -4\gamma_1 \gamma_2 A \{(\gamma_1 + \gamma_2) \sin 2\theta - \sin 4\theta\} \\ &\quad + B \{[3 - (\gamma_1 + \gamma_2)^2 - \gamma_1^2 \gamma_2^2 - 3\gamma_1 \gamma_2] \sin 2\theta \\ &\quad + (\gamma_1 + \gamma_2)(1 + \gamma_1 \gamma_2) \sin 4\theta - \gamma_1 \gamma_2 \sin 6\theta\}. \end{aligned} \right\} \quad (12.3)$$

For spruce $B = 0.527G/(2\pi a^2)$ and the numerical values of the edge stresses are given in table 3 where the line $\theta = 0$ is chosen to be parallel to the grain. The values of $\hat{\theta}\hat{\theta}$, $\hat{r}\hat{\theta}$ and $\hat{r}r$ are shown in figure 2 together with the values of $\hat{\theta}\hat{\theta}$ when a uniform couple G is applied at the edge of the hole.

BOLT IN A TENSION MEMBER

13. A problem of some practical importance arises when a pin or bolt is inserted into a circular hole in a piece of timber under tension, the natural diameter of the bolt being greater than that of the hole, and a solution of the problem is given in this section. Coker and Filon examined the problem when the materials of the bolt and plate are both isotropic by extending some work of Suyehiro (1914). They assumed that

- (i) the friction between the pin and the plate could be neglected;
- (ii) contact is maintained between the bolt and the plate throughout.

The same assumptions are made here so that at the boundary $r = a$,

$$\hat{r}r_1 = \hat{r}r_2, \quad (13.1)$$

$$\hat{r}\hat{\theta}_1 = \hat{r}\hat{\theta}_2 = 0, \quad (13.2)$$

$$(u_r)_2 = (u_r)_1 + c, \quad (13.3)$$

the suffixes 1 and 2 referring to the bolt and the plate respectively, and c being the excess of the natural radius of the bolt over the natural radius of

TABLE 3. VALUES OF STRESSES AT THE EDGE OF THE CIRCLE; SPRUCE

θ°	$\frac{2\pi a^2 \widehat{rr}}{G}$	$\frac{2\pi a^2 \widehat{\theta\theta}}{G}$	$\frac{2\pi a^2 \widehat{r\theta}}{G}$	$\frac{2\pi a^2 \widehat{xx}}{G}$	$\frac{2\pi a^2 \widehat{yy}}{G}$	$\frac{2\pi a^2 \widehat{xy}}{G}$
0	0	0	0.473	0	0	-0.473
10	0.180	0.117	0.504	0.351	-0.054	-0.463
20	0.339	0.269	0.596	0.714	-0.106	-0.434
30	0.457	0.491	0.736	1.10	-0.155	-0.383
40	0.519	0.818	0.908	1.54	-0.200	-0.305
50	0.519	1.28	1.09	2.04	-0.241	-0.185
60	0.457	1.93	1.26	2.66	-0.269	-0.006
70	0.339	2.71*	1.40	3.34	-0.286	0.314
75	0.264	3.01	1.46	3.55	-0.281	0.575
80	0.180	2.92	1.50	3.35	-0.249	0.937
85	0.092	1.98	1.52	2.23	-0.158	1.33
90	0	0	1.53	0	0	1.53

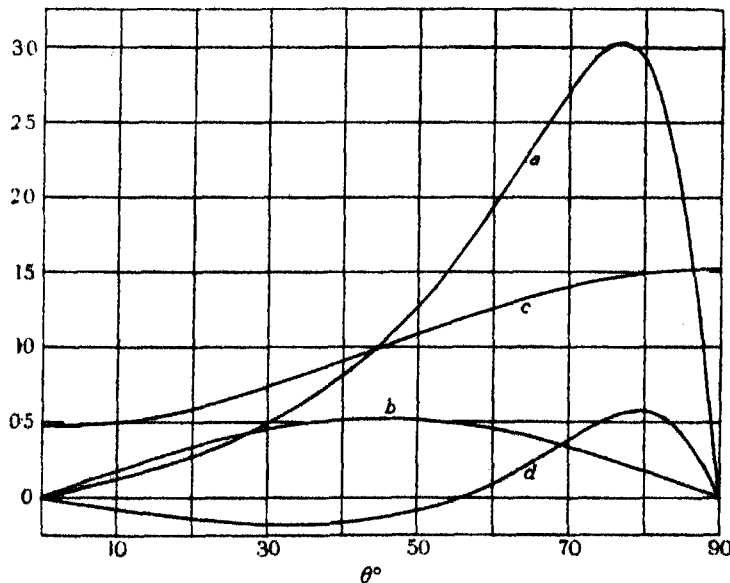


FIGURE 2. Stress distribution at the edge of the circle $r=a$ in spruce. Grain parallel to $\theta=0$. Curve (a) $2\pi a^2 \widehat{\theta\theta}/G$. Curve (b) $2\pi a^2 \widehat{rr}/G$. Curve (c) $2\pi a^2 \widehat{r\theta}/G$. Curve (d) $2\pi a^2 \widehat{\theta\theta}/G$ for uniform couple at the hole.

the hole. The plate is aeolotropic of the type considered throughout this paper and the bolt will be supposed to be isotropic with Young's modulus E' and Poisson's ratio σ' . The stress functions are then taken to be

$$\chi = C_0 r^2 + \sum_{n=1}^{\infty} C_{2n} \{ (2n+1) a^2 r^{2n} \cos 2n\theta - (2n-1) r^{2n+2} \cos 2n\theta \} / a^{2n} \quad (r \leq a), \quad (13.4)$$

$$\text{and} \quad \chi = \frac{1}{2} T y^2 + A_0 \chi_0 + A_2 \chi_2 + B_2 \chi'_2 + \sum_{n=2}^{\infty} A_{2n} (\chi_{2n} + \chi'_{2n}) \quad (r \geq a). \quad (13.5)$$

These give a tension T at infinity parallel to the x -axis and single-valued stresses and displacements at all points of the material and the bolt. Condition (13.2) gives

$$A_2 - B_2 = \frac{1}{2} T, \quad (13.6)$$

and condition (13.1) requires that

$$\left. \begin{aligned} A_0 + \frac{1}{2} T &= 2C_0, & A_2 + B_2 + \frac{1}{2} T &= -6C_2, \\ A_{2n} &= -(4n^2 - 1) C_{2n} \quad (n \geq 2). \end{aligned} \right\} \quad (13.7)$$

After some straightforward calculation it is found that the condition (13.3) is satisfied if

$$\left. \begin{aligned} (\rho_2 + \rho_4) A_0 + (\rho_1 - \rho_3) B_2 + (s_{11} + s_{12}) T &= 4(1 - \sigma') C_0 / E' + 2c/a, \\ (\rho_4 - \rho_2) A_0 + \frac{1}{2} (\rho_2 + \rho_4) A_2 - (\rho_1 + \rho_3) B_2 + \frac{1}{2} (\rho_1 - \rho_3) A_4 + (s_{11} - s_{12}) T \\ &= -4(3 + \sigma') C_2 / E', \\ \frac{(\rho_1 - \rho_3) A_{2n+2}}{2n+1} + \left(\frac{\rho_2 + \rho_4}{2n+1} - \frac{\rho_1 + \rho_3}{2n-1} \right) A_{2n} + \frac{(\rho_4 - \rho_2) A_{2n-2}}{2n-1} \\ &= \frac{4(1 - 4n - \sigma') C_{2n}}{E'} \quad (n \geq 2). \end{aligned} \right\} \quad (13.8)$$

When the coefficients in the above equations have been evaluated the stresses in the plate at the edge of the circle $r = a$ are given by

$$\bar{r}_2 = \frac{1}{2} T (1 + \cos 2\theta) + A_0 + (A_2 + B_2) \cos 2\theta + 2 \sum_{n=2}^{\infty} A_{2n} \cos 2n\theta, \quad (13.9)$$

$$\bar{\theta}_2 = \frac{1}{2} T (1 - \cos 2\theta) + \frac{\sum_{n=0}^{\infty} K_{2n} \cos 2n\theta}{(1 + \gamma_1^2 - 2\gamma_1 \cos 2\theta)(1 + \gamma_2^2 - 2\gamma_2 \cos 2\theta)}, \quad (13.10)$$

where

$$K_0 = \{3\gamma_1^2\gamma_2^2 - (\gamma_1 + \gamma_2)^2 - 1\} A_0 + (\gamma_1 + \gamma_2)(1 + \gamma_1\gamma_2) A_2 \\ + (\gamma_1 + \gamma_2)(\gamma_1\gamma_2 - 3) B_2 + 2\gamma_1\gamma_2 A_4, \quad (13.11)$$

$$K_2 = 2(\gamma_1 + \gamma_2)(1 - \gamma_1\gamma_2) A_0 + \{3\gamma_1^2\gamma_2^2 - \gamma_1\gamma_2 - (\gamma_1 + \gamma_2)^2 - 1\} A_2 \\ + \{3 - (\gamma_1 + \gamma_2)^2 + 3\gamma_1\gamma_2 - \gamma_1^2\gamma_2^2\} B_2 \\ + 2(\gamma_1 + \gamma_2)(\gamma_1\gamma_2 - 1) A_4 + 2\gamma_1\gamma_2 A_6, \quad (13.12)$$

$$K_4 = 2\gamma_1\gamma_2 A_0 + (\gamma_1 + \gamma_2)(1 - 3\gamma_1\gamma_2) A_2 + (\gamma_1 + \gamma_2)(1 + \gamma_1\gamma_2) B_2 \\ + 2\{1 - (\gamma_1 + \gamma_2)^2 + \gamma_1^2\gamma_2^2\} A_4 \\ + 2(\gamma_1 + \gamma_2)(\gamma_1\gamma_2 - 1) A_6 + 2\gamma_1\gamma_2 A_8, \quad (13.13)$$

$$K_6 = 3\gamma_1\gamma_2 A_2 - \gamma_1\gamma_2 B_2 + 2(\gamma_1 + \gamma_2)(1 - \gamma_1\gamma_2)(A_4 - A_8) \\ + 2\{1 - (\gamma_1 + \gamma_2)^2 + \gamma_1^2\gamma_2^2\} A_6 + 2\gamma_1\gamma_2 A_{10}, \quad (13.14)$$

$$K_{2n} = 2\{1 - (\gamma_1 + \gamma_2)^2 + \gamma_1^2\gamma_2^2\} A_{2n} + 2(\gamma_1 + \gamma_2)(1 - \gamma_1\gamma_2)(A_{2n-2} - A_{2n+2}) \\ + 2\gamma_1\gamma_2(A_{2n-4} + A_{2n+4}) \quad (n \geq 4). \quad (13.15)$$

If the material of the plate is isotropic all the coefficients are zero except A_0 , A_2 , B_2 , C_0 , C_2 and the solution is then given in a finite number of terms as shown by Coker & Filon.

If the bolt is to remain in contact with the plate at all points of the circle $r = a$ then the normal stress must always be a pressure and this means that $c/(aT)$ must exceed a certain number which will be denoted by λ . If T exceeds $c/(a\lambda)$ a gap will appear between the bolt and the plank at the points where the diameter parallel to the line of the tension cuts the circular boundary. It is of some interest to examine the case when contact just ceases at these points, i.e. when $T = c/(a\lambda)$, and calculations have been made for a bolt in a spruce plank. Calculations for an isotropic plank have also been made for comparison.

The elastic constants of the bolt may take various values so for simplicity the calculations have been confined to the case when the material of the bolt is supposed to be very rigid compared with the material of the plank, in which case the C terms of the right-hand side of equations (13.8) may be neglected. The extra calculation which is required to allow for the elasticity of the bolt is not great.

For a rigid bolt in an isotropic plate $E\lambda = (11 + 5\sigma)/(5 - \sigma)$ and when $T = c/(a\lambda)$ the edge stresses in the plate at the boundary reduce to

$$\bar{r}_2 = -6T(1 - \cos 2\theta)/(5 - \sigma), \quad \bar{\theta}\theta_2 = T\{11 - \sigma - 2(2 - \sigma)\cos 2\theta\}/(5 - \sigma). \quad (13.16)$$

TABLE 4. VALUES OF COEFFICIENTS FOR SPRUCE

Tension parallel to grain							
λ/s_{22}	B_2/T	A_0/T	A_2/T	A_4/T	A_6/T	A_8/T	A_{10}/T
0.208	-0.408	-0.824	0.0924	0.0404	0.0171	0.0071	0.0029
A_{12}/T	A_{14}/T	A_{16}/T	A_{18}/T	K_0/T	K_2/T	K_4/T	
0.0012	0.0005	0.0002	0.0001	0.313	-0.0764	0.212	
K_6/T	K_8/T	K_{10}/T	K_{12}/T	K_{14}/T	K_{16}/T	K_{18}/T	
0.0185	-0.0065	-0.0027	-0.0011	-0.0005	-0.0002	-0.0001	
Tension perpendicular to grain							
λ/s_{22}	B_2/T	A_0/T	A_2/T	A_4/T	A_6/T	A_8/T	A_{10}/T
44.2	3.19	-5.62	3.69	-1.62	0.684	-0.284	0.116
A_{12}/T	A_{14}/T	A_{16}/T	A_{18}/T	A_{20}/T	K_0/T	K_2/T	K_4/T
-0.0473	0.0191	-0.0077	0.0030	-0.0011	4.33	-2.52	1.28
K_6/T	K_8/T	K_{10}/T	K_{12}/T	K_{14}/T	K_{16}/T	K_{18}/T	K_{20}/T
-0.574	0.260	-0.109	0.0451	-0.0185	0.0076	-0.0034	0.0020

TABLE 5. VALUES OF STRESSES ON THE EDGE OF THE HOLE
FILLED WITH A SMOOTH RIGID BOLT: SPRUCE

θ°	\widehat{rr}/T	$\widehat{\theta\theta}/T$	\widehat{xx}/T	\widehat{yy}/T	\widehat{xy}/T
Tension parallel to grain					
0	0	0.143	0	0.143	0
10	-0.072	0.165	-0.065	0.158	-0.041
20	-0.205	0.224	-0.155	0.174	-0.138
30	-0.308	0.318	-0.152	0.162	-0.271
40	-0.371	0.465	-0.028	0.119	-0.411
50	-0.409	0.705	0.245	0.051	-0.549
60	-0.430	1.140	0.748	-0.038	-0.680
70	-0.443	2.03	1.74	-0.154	-0.794
75	-0.447	2.81	2.60	-0.229	-0.815
80	-0.450	3.91	3.78	-0.318	-0.746
85	-0.451	5.16	5.11	-0.408	-0.487
90	-0.452	5.79	5.79	-0.452	0
Tension perpendicular to grain					
0	0	22.3	0	22.3	0
5	-0.019	19.8	0.131	19.7	-1.72
10	-0.080	14.9	0.371	14.4	-2.56
15	-0.187	10.6	0.534	9.86	-2.69
20	-0.333	7.56	0.590	6.64	-2.54
30	-0.848	4.40	0.465	3.09	-2.27
40	-1.72	3.15	0.290	1.14	-2.40
50	-3.22	2.71	0.259	-0.770	-2.92
60	-5.76	2.71	0.589	-3.64	-3.67
70	-9.87	3.03	1.52	-8.36	-4.15
80	-15.2	3.56	3.00	-14.6	-3.21
90	-18.1	3.87	3.87	-18.1	0

The solution does not reduce to a finite number of terms when the material of the plank is not isotropic. The coefficients A_0, A_2, \dots for spruce have been found from equations (13.6) and (13.8) by neglecting every coefficient

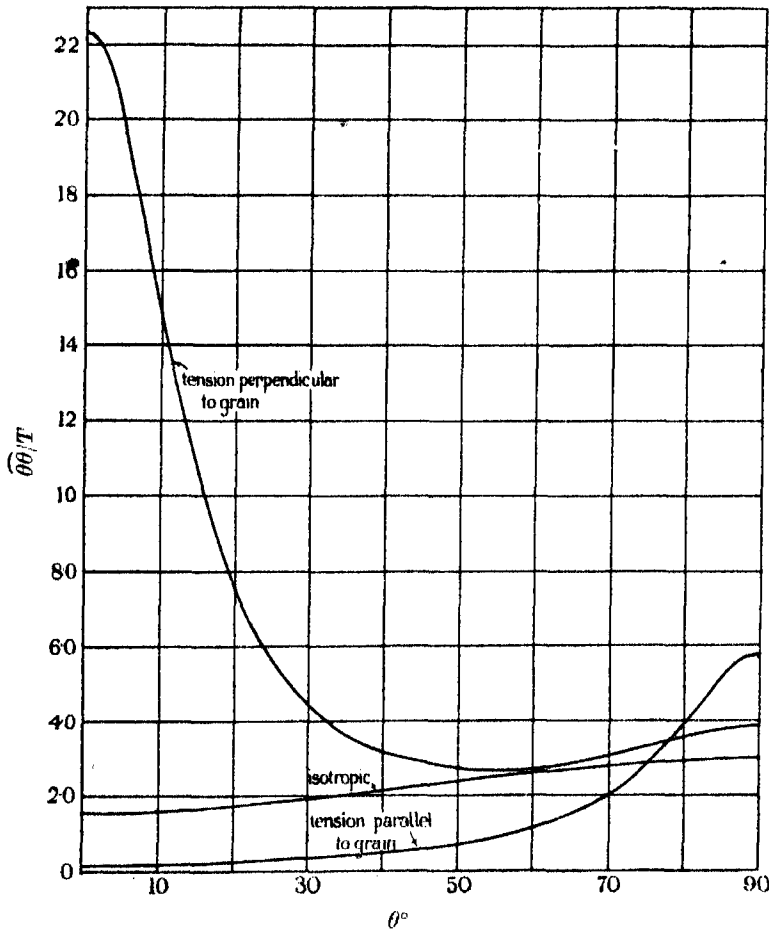


FIGURE 3. Values of circumferential stress at the edge of a smooth bolt. Tension parallel to $\theta = 0$.

beyond A_{30} as it is found that this gives sufficient accuracy. When $T = c/(a\lambda)$ the values of λ and the coefficients are recorded in table 4 for the cases when the tension is parallel and perpendicular to the grain. The corresponding values of the stresses at the edge of the hole are given in table 5 where $\theta = 0$ is a line through the centre of the circular bolt, parallel to the applied

tension. The values of $\theta\theta$ and $\hat{r}r$ are shown in figures 3 and 4 respectively together with the corresponding values for an isotropic material whose Poisson's ratio is $\frac{1}{4}$.

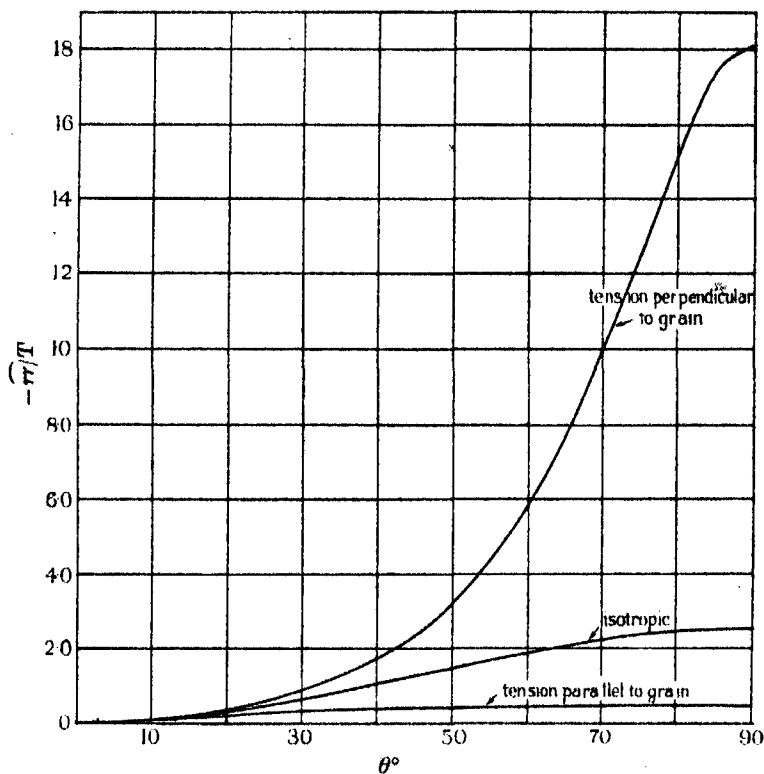


FIGURE 4. Values of radial stress at the edge of a smooth bolt.
Tension parallel to $\theta = 0$.

BOLTS OR KNOTS IN A TENSION MEMBER

14. The condition that friction between the bolt and the plate is negligible is probably not accurately fulfilled in practice. Any practical estimate of the effect of friction seems to be difficult, but it is of some interest to consider the limiting case in which the surfaces are perfectly rough so that the conditions at $r = a$ become

$$\hat{r}r_1 = \hat{r}r_2, \quad (14.1)$$

$$r\theta_1 = r\theta_2, \quad (14.2)$$

$$(u_r)_2 = (u_r)_1 + c, \quad (u_\theta)_2 = (u_\theta)_1. \quad (14.3)$$

When $c = 0$ these conditions will also be applicable to the problem of a circular discontinuity (such as a live knot) in a tension member if the discontinuity is rigidly connected with the rest of the plate. It is found that a solution of the tension problem satisfying conditions (14.1)–(14.3) can be expressed in finite terms even if the material of the bolt or discontinuity is aeolotropic as well as that of the plank. If the material of the bolt or discontinuity is aeolotropic with two directions of symmetry in the same directions as those of the plate, suitable stress functions are

$$\chi = C\Psi_0 + D\Psi_2 \quad (r \leq a), \quad (14.4)$$

and

$$\chi = \frac{1}{2}Ty^2 + A\chi_0 + B\chi_2' \quad (r \geq a), \quad (14.5)$$

if there is a uniform tension T at infinity parallel to the x -axis. The conditions (14.1) and (14.2) give

$$A + \frac{1}{2}T = C, \quad B + \frac{1}{2}T = D, \quad (14.6)$$

and (14.3) can be satisfied if

$$\left. \begin{aligned} a\rho_4 A - a\rho_3 B + as_{11} T &= a(s'_{11} + s'_{12}) C + a(s'_{11} - s'_{12}) D + c, \\ a\rho_2 A + a\rho_1 B + as_{12} T &= a(s'_{22} + s'_{12}) C + a(s'_{12} - s'_{22}) D + c. \end{aligned} \right\} \quad (14.7)$$

The stresses in the plank at the edge of the circle $r = a$ are

$$\widehat{rr}_2 = A + \frac{1}{2}T + (B + \frac{1}{2}T) \cos 2\theta, \quad \widehat{r\theta}_2 = -(B + \frac{1}{2}T) \sin 2\theta, \quad (14.8)$$

$$\widehat{\theta\theta}_2 = \frac{1}{2}T(1 - \cos 2\theta) + \frac{\sum_{n=0}^3 K_{2n} \cos 2n\theta}{(1 + \gamma_1^2 - 2\gamma_1 \cos 2\theta)(1 + \gamma_2^2 - 2\gamma_2 \cos 2\theta)}, \quad (14.9)$$

where

$$K_0 = \{3\gamma_1^2\gamma_2^2 - (\gamma_1 + \gamma_2)^2 - 1\} A + (\gamma_1 + \gamma_2)(\gamma_1\gamma_2 - 3) B, \quad (14.10)$$

$$K_2 = 2(\gamma_1 + \gamma_2)(1 - \gamma_1\gamma_2) A + \{3 - (\gamma_1 + \gamma_2)^2 - \gamma_1^2\gamma_2^2 + 3\gamma_1\gamma_2\} B, \quad (14.11)$$

$$K_4 = 2\gamma_1\gamma_2 A + (\gamma_1 + \gamma_2)(1 + \gamma_1\gamma_2) B, \quad K_6 = -\gamma_1\gamma_2 B. \quad (14.12)$$

Consider first the problem of a rough bolt in a tension member. As in § 13 the bolt will remain in contact with the plank provided $c/(aT)$ exceeds a certain number λ and calculations are again carried out for the case when $T = c/(a\lambda)$, i.e. when the bolt is only just in contact with the plank at the end of the diameter of the circle which is parallel to the applied tension. Also, calculations are again confined to a rigid bolt. For an isotropic plank $E\lambda = (5 + \sigma)/(3 - \sigma)$ and the edge stresses for the case $T = c/(a\lambda)$ are

$$\left. \begin{aligned} \widehat{rr} &= -\frac{2T}{3 - \sigma}(1 - \cos 2\theta), \quad \widehat{r\theta} = -\frac{2T}{3 - \sigma} \sin 2\theta, \\ \widehat{\theta\theta} &= \frac{T}{3 - \sigma} \{5 - \sigma + 2\sigma \cos 2\theta\}. \end{aligned} \right\} \quad (14.13)$$

The values of λ , A , B for spruce are given in table 6 and the corresponding stresses at the edge of the circle are recorded in table 7, where $\theta = 0$ is parallel to the applied tension. Values of $\bar{\theta}\theta$ and $\bar{r}r$ are shown in figures 5 and 6 respectively.

TABLE 6. VALUES OF COEFFICIENTS FOR ROUGH BOLT IN SPRUCE

tension parallel to grain			tension perpendicular to grain		
λ/s_{22}	A/T	B/T	λ/s_{22}	A/T	B/T
0.179	-0.679	-0.321	26.7	-3.47	2.47

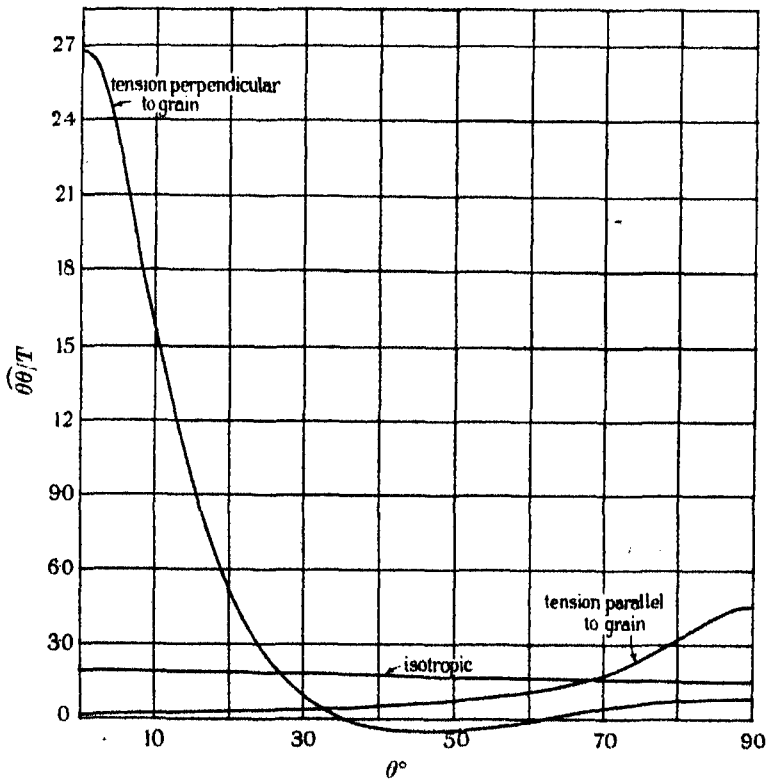


FIGURE 5. Values of circumferential stress at the edge of a rough bolt. Tension parallel to $\theta = 0$.

When the tension is applied parallel to the grain in spruce the shear stress $\bar{r}\theta$ at the edge of a rough bolt is quite small and the values of $\bar{\theta}\theta$ and $\bar{r}r$ at the edge of a rough bolt do not differ greatly from those at the edge of a

TABLE 7. VALUES OF STRESSES ON THE EDGE OF THE HOLE
FILLED WITH A ROUGH RIGID BOLT: SPRUCE

θ°	\widehat{rr}/T	$\widehat{\theta\theta}/T$	$\widehat{r\theta}/T$	\widehat{xx}/T	\widehat{yy}/T	\widehat{xy}/T
Tension parallel to grain						
0	0	0.179	0	0	0.179	0
10	-0.011	0.200	-0.061	0.016	0.172	-0.094
20	-0.042	0.263	-0.115	0.068	0.154	-0.186
30	-0.089	0.371	-0.155	0.160	0.122	-0.277
40	-0.148	0.531	-0.176	0.306	0.077	-0.365
50	-0.210	0.764	-0.176	0.535	0.019	-0.449
60	-0.268	1.13	-0.155	0.916	-0.053	-0.529
70	-0.316	1.81	-0.115	1.64	-0.141	-0.596
75	-0.334	2.39	-0.089	2.25	-0.196	-0.604
80	-0.347	3.19	-0.061	3.10	-0.262	-0.547
85	-0.355	4.08	-0.031	4.05	-0.327	-0.354
90	-0.358	4.53	0	4.53	-0.358	0
Tension perpendicular to grain						
0	0	26.8	0	0	26.8	0
5	-0.045	23.1	-0.516	0.220	22.8	-2.51
10	-0.179	15.8	-1.02	0.649	14.9	-3.68
15	-0.398	9.44	-1.49	1.00	8.04	-3.75
20	-0.695	5.12	-1.91	1.21	3.21	-3.33
30	-1.49	0.892	-2.57	1.34	-1.93	-2.32
40	-2.46	-0.394	-2.93	1.28	-4.13	-1.52
50	-3.49	-0.477	-2.93	1.16	-5.13	-0.975
60	-4.46	-0.088	-2.57	1.05	-5.60	-0.606
70	-5.25	0.392	-1.91	0.960	-5.82	-0.349
80	-5.77	0.752	-1.02	0.903	-5.92	-0.159
90	-5.95	0.884	0	0.884	-5.95	0

smooth bolt, except in small regions at the ends of the diameter of the bolt which is perpendicular to the grain. The differences between rough and smooth bolts is rather more marked when the tension is applied perpendicular to the grain in spruce, also for isotropic materials, and the stresses are compared graphically for an isotropic material in figure 7.

The problem of a live knot in a tension member may also be discussed with the help of the analysis of this section by putting $c = 0$. The solution then assumes that the knot is connected continuously with the rest of the plank. The actual values of the elastic constants of the knot will be very difficult to determine, but the limiting case in which the knot is supposed to be very rigid compared with the rest of the plank is of some interest and calculations have been made for this case. The problem is then really equivalent to a circular hole with clamped edges in a tension member. The

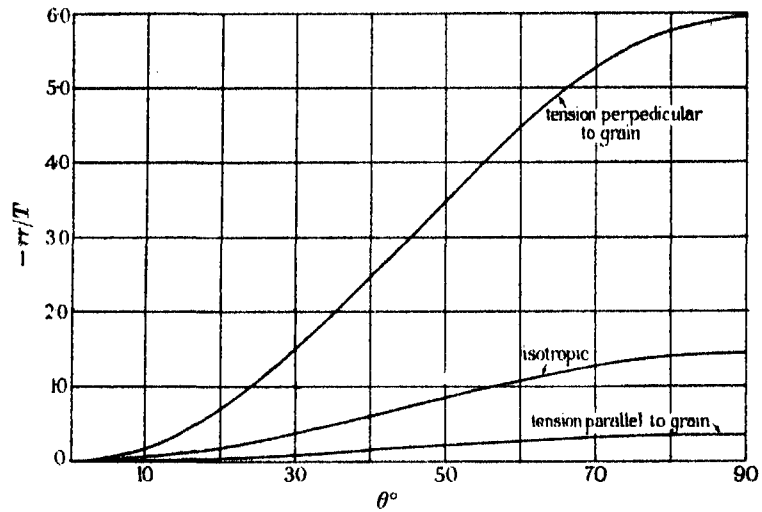


FIGURE 6. Values of radial stress at the edge of a rough bolt.
Tension parallel to $\theta = 0$.

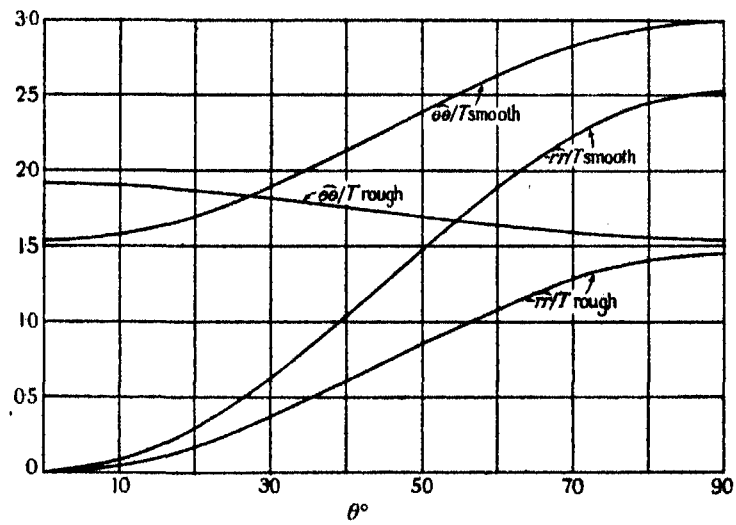


FIGURE 7. Stresses at the edge of a smooth and rough bolt
in an isotropic tension member.

values of the coefficients A, B are given in table 8 and the corresponding edge stresses are recorded in table 9, where $\theta = 0$ is parallel to the tension. The distributions of $\bar{\theta}\bar{\theta}$ and $\bar{r}\bar{r}$ are shown in figure 8 for the case when the tension is parallel to the grain and in figure 9 for the case when the tension is perpendicular to the grain. The values of $\bar{\theta}\bar{\theta}$ around an unstressed circular hole in a tension member are also shown in figures 8 and 9.

TABLE 8. VALUES OF COEFFICIENTS FOR LIVE KNOT IN SPRUCE

tension parallel to grain		tension perpendicular to grain	
A/T	B/T	A/T	B/T
0.105	0.0924	0.963	0.133

TABLE 9. VALUES OF STRESSES AT THE EDGE OF A KNOT IN A SPRUCE TENSION MEMBER

θ°	$\bar{r}\bar{r}/T$	$\bar{\theta}\bar{\theta}/T$	$\bar{r}\bar{\theta}/T$	$\bar{x}\bar{x}/T$	$\bar{y}\bar{y}/T$	$\bar{x}\bar{y}/T$
Tension parallel to grain						
0	1.20	0.026	0	1.20	0.026	0
10	1.16	0.060	-0.203	1.20	0.024	-0.002
20	1.06	0.158	-0.381	1.20	0.019	-0.002
30	0.901	0.307	-0.513	1.20	0.011	0.001
40	0.708	0.481	-0.583	1.19	0.000	0.010
50	0.502	0.648	-0.583	1.16	-0.012	0.029
60	0.309	0.759	-0.513	1.09	-0.023	0.062
70	0.151	0.727	-0.381	0.904	-0.026	0.107
75	0.092	0.609	-0.296	0.723	-0.022	0.127
80	0.048	0.401	-0.203	0.460	-0.011	0.130
85	0.021	0.141	-0.103	0.158	0.004	0.091
90	0.012	0.005	0	0.005	0.012	0
Tension perpendicular to grain						
0	2.10	1.18	0	2.10	1.18	0
5	2.09	0.777	-0.110	2.10	0.768	0.005
10	2.06	-0.005	-0.217	2.07	-0.017	0.149
15	2.01	-0.647	-0.317	1.99	-0.627	0.390
20	1.95	-1.02	-0.407	1.86	-0.938	0.643
30	1.78	-1.22	-0.548	1.50	-0.946	1.03
40	1.57	-1.05	-0.624	1.10	-0.582	1.18
50	1.35	-0.760	-0.624	0.727	-0.134	1.15
60	1.15	-0.455	-0.548	0.420	0.271	0.967
70	0.978	-0.203	-0.407	0.197	0.578	0.691
80	0.868	-0.039	-0.217	0.062	0.766	0.359
90	0.829	0.017	0	0.017	0.829	0

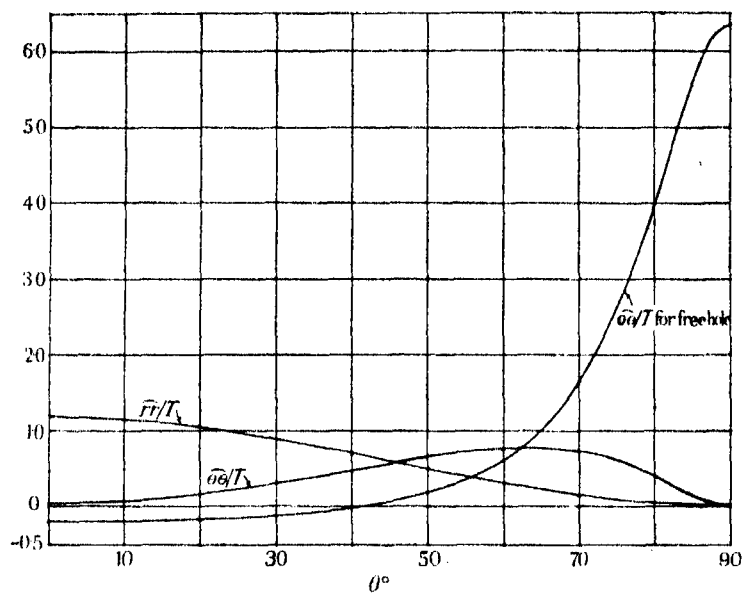


FIGURE 8. Stresses at the edge of a knot in a spruce plank.
Tension parallel to the grain and to $\theta=0$.

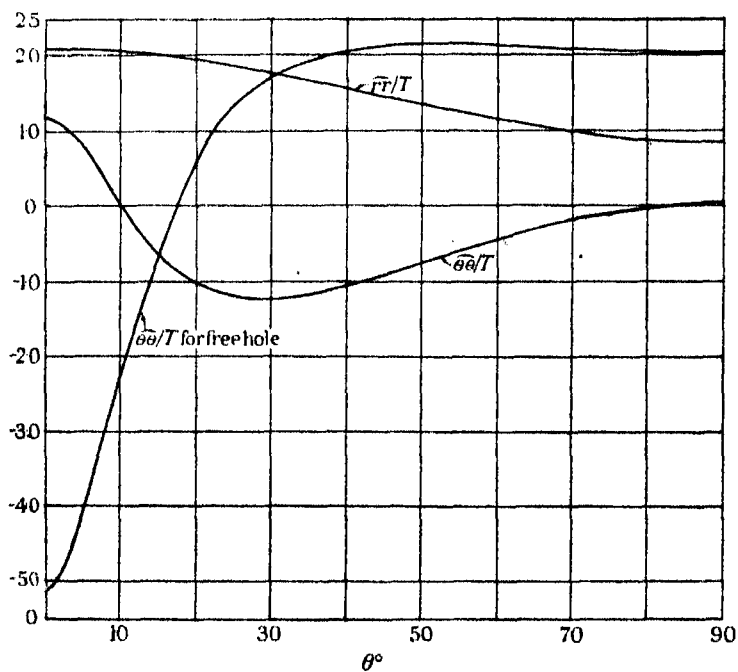


FIGURE 9. Stresses at the edge of a knot in a spruce plank.
Tension parallel to $\theta=0$ and perpendicular to the grain.

When the problem of an unstressed circular hole in a spruce tension member was solved (Green & Taylor 1945) an estimate of the type of failure to be expected was made by applying ultimate yield values for spruce in conjunction with the calculated stresses. A similar process can be carried out for the problem of a knot in a tension member and it is found that the results are very similar in character to those found for a hole. In particular, when the tension is applied parallel to the grain failure may take place at the edge of the knot by a breakdown in shear parallel to the grain, the shear cracks being in almost the same positions as they are for the hole.

KNOTS IN PLANK BENT BY COUPLES

15. In this section the plank is supposed to be bounded by two straight edges $y = \pm b$, where b is so large that the edges may be thought of as being at infinite distance apart. The plank is acted on by a couple so that there is a pure bending moment of amount M across every section of the plate parallel to the y -axis. The plank contains a circular knot of radius a whose centre is at the origin of co-ordinates. As attention will be confined to a knot which is supposed to be very rigid compared with the rest of the plank it is sufficient to consider conditions at the edge of the knot which are equivalent to clamping. A suitable stress function for the plank is then

$$\chi = \frac{1}{4}My^3/b^3 + A\Phi_1 + B\Phi_3, \quad (15.1)$$

and the conditions at the edge of the knot require that

$$\left. \begin{aligned} \rho_4 A - \rho_3 B &= -\frac{3}{2}s_{11}Ma/b^3, \\ \rho_2 A + \rho_1 B &= -\frac{3}{4}(s_{11} + s_{12})Ma/b^3. \end{aligned} \right\} \quad (15.2)$$

The edge stresses at $r = a$ reduce to

$$\bar{r}\bar{r} = (A + 3Ma/8b^3)\sin\theta + (B + 3Ma/8b^3)\sin 3\theta, \quad (15.3)$$

$$\bar{r}\bar{\theta} = -(A + 3Ma/8b^3)\cos\theta + (B + 3Ma/8b^3)\cos 3\theta, \quad (15.4)$$

$$\bar{\theta}\bar{\theta} = \frac{3Ma}{8b^3}(3\sin\theta - \sin 3\theta) + \frac{\sum_{n=0}^3 K_{2n+1}\sin(2n+1)\theta}{(1 + \gamma_1^2 - 2\gamma_1\cos 2\theta)(1 + \gamma_2^2 - 2\gamma_2\cos 2\theta)}, \quad (15.5)$$

where

$$K_1 = \{3\gamma_1^2\gamma_2^2 - (\gamma_1 + \gamma_2)^2 - 1 - (\gamma_1 + \gamma_2)(1 + \gamma_1\gamma_2)\}A + \{(\gamma_1 + \gamma_2)(\gamma_1\gamma_2 - 3) - 3\gamma_1\gamma_2\}B, \quad (15.6)$$

$$K_3 = \{(\gamma_1 + \gamma_2)(1 - 3\gamma_1\gamma_2) + \gamma_1\gamma_2\}A + \{3 - (\gamma_1 + \gamma_2)^2 - \gamma_1^2\gamma_2^2\}B, \quad (15.7)$$

$$K_5 = 3\gamma_1\gamma_2A + (\gamma_1 + \gamma_2)(1 + \gamma_1\gamma_2)B, \quad K_7 = -\gamma_1\gamma_2B. \quad (15.8)$$

TABLE 10. VALUES OF COEFFICIENTS FOR LIVE KNOT
IN SPRUCE PLANK BENT BY COUPLES

couple across sections perpendicular to grain		couple across sections parallel to grain	
A	0.203	A	4.01
B	0.135	B	-1.63
K_1	0.108	K_1	-5.33
K_3	0.232	K_3	-1.48
K_5	-0.0626	K_5	-0.444
K_7	-0.0091	K_7	0.110

TABLE 11. VALUES OF STRESSES AT THE EDGE OF A KNOT
IN A SPRUCE PLANK BENT BY COUPLES

θ°	$\frac{b^2 \widehat{rr}}{aM}$	$\frac{b^2 \widehat{\theta\theta}}{aM}$	$\frac{b^2 \widehat{r\theta}}{aM}$	$\frac{b^2 \widehat{xx}}{aM}$	$\frac{b^2 \widehat{yy}}{aM}$	$\frac{b^2 \widehat{xy}}{aM}$
Grain parallel to line $\theta = 0^\circ$						
0	0	0	-0.068	0	0	-0.068
10	0.355	0.033	-0.128	0.389	-0.0006	-0.065
20	0.639	0.123	-0.288	0.764	-0.0023	-0.055
30	0.799	0.304	-0.501	1.11	-0.0058	-0.036
40	0.813	0.579	-0.698	1.40	-0.0113	-0.006
50	0.698	0.904	-0.813	1.62	-0.0181	0.040
60	0.501	1.18	-0.799	1.70	-0.0221	0.106
70	0.288	1.22	-0.639	1.52	-0.0142	0.192
75	0.198	1.05	-0.510	1.25	-0.0002	0.229
80	0.128	0.712	-0.355	0.816	0.0238	0.234
85	0.083	0.270	-0.182	0.300	0.0531	0.163
90	0.068	0.035	0	0.035	0.068	0
Grain perpendicular to line $\theta = 0^\circ$						
0	0	0	-5.64	0	0	-5.64
5	0.056	-7.12	-5.58	0.971	-8.04	-4.88
10	0.132	-10.4	-5.41	1.66	-12.0	-3.28
15	0.244	-10.6	-5.13	2.08	-12.4	-1.73
20	0.409	-9.35	-4.75	2.32	-11.3	-0.503
30	0.933	-6.25	-3.80	2.43	-7.74	1.21
40	1.73	-3.73	-2.73	2.16	-4.16	2.21
50	2.73	-1.99	-1.73	1.66	-0.922	2.62
60	3.80	-0.887	-0.933	1.09	1.82	2.50
70	4.75	-0.264	-0.409	0.585	3.90	1.92
80	5.41	0.034	-0.132	0.241	5.20	1.04
90	5.64	0.120	0	0.120	5.64	0

For an isotropic material the stresses at the edge of a rigid knot reduce to

$$\hat{r}r = \hat{\theta}\theta/\sigma = \frac{3Ma}{2b^3} \left(\frac{\sin \theta}{1+\sigma} + \frac{\sin 3\theta}{3-\sigma} \right), \quad \hat{r}\theta = \frac{3Ma}{2b^3} \left(-\frac{\cos \theta}{1+\sigma} + \frac{\cos 3\theta}{3-\sigma} \right). \quad (15.9)$$

Calculations have been carried out for an isotropic material whose Poisson's ratio is $\frac{1}{4}$ and also for a spruce plank. The values of the constants A, \dots, K_7 for spruce are given in table 10, the corresponding stresses being

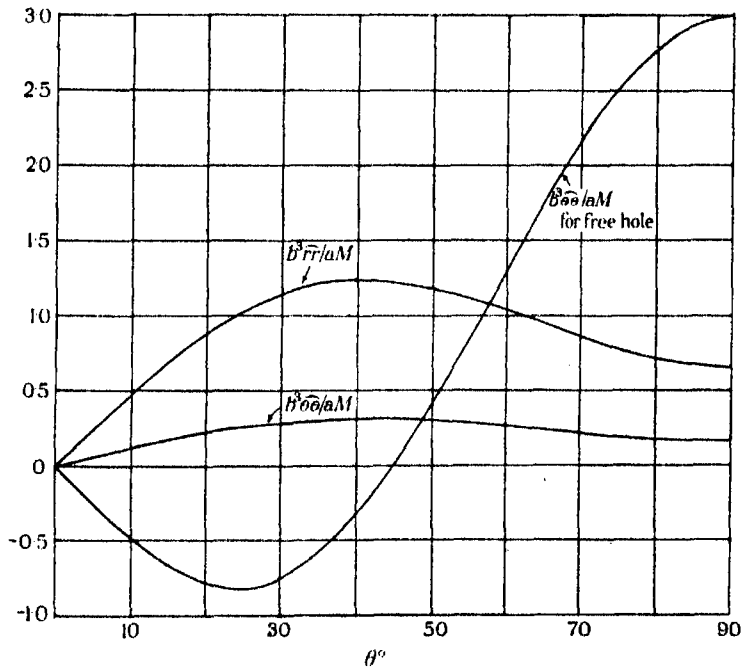


FIGURE 10. Stress distribution at the edge of a circular discontinuity in an isotropic plate bent by couples.

recorded in table 11 where the bending moment acts across sections which are perpendicular to $\theta = 0$. The distribution of stress is shown graphically in figures 10, 11 and 12 respectively for an isotropic material, for spruce when the bending moment is across sections which are perpendicular to the grain and for spruce when the bending moment is across sections which are parallel to the grain. The corresponding values of the stress at the edge of a free hole are included for comparison. In table 10 each coefficient is to be multiplied by aM/b^3 .

Bending tests on wooden planks are usually carried out on planks cut in the L, T or L, R planes and bent by couples across sections which are

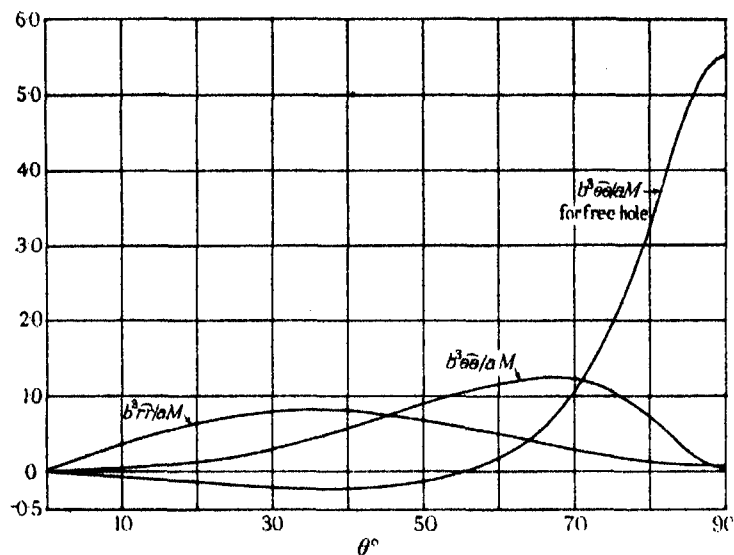


FIGURE 11. Stresses at the edge of a knot in a spruce plank. Bending moment applied across sections which are perpendicular to the grain and to $\theta = 0$.

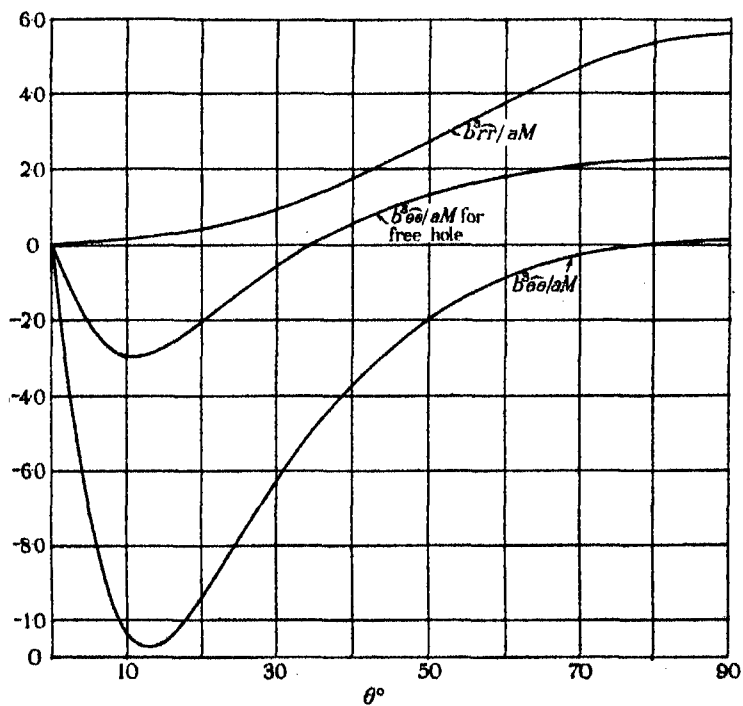


FIGURE 12. Stresses at the edge of a knot in a spruce plank. Bending moment applied across sections which are parallel to the grain and perpendicular to $\theta = 0$.

perpendicular to the L direction, and failure would normally be expected by breakdown in fibres in compression at the edge of the plank. When the plank contains no knots or holes the maximum value of $b^2\hat{x}/M$ at the edge of the plank is 1.5. When an isotropic plank bent by couples is pierced by a central hole Howland & Stevenson (1933) have shown that the maximum stress at the edge of the plank is only very slightly less than 1.5 even when the hole occupies half the width of the plank. The same result might be expected for a wooden plank when the L direction is parallel to the edge. It is of interest to enquire whether a central knot or hole in a wooden plank is likely to affect the kind of failure to be expected in bending tests. By applying values of ultimate yield stresses which have been used previously (Green & Taylor 1945) to the specimen of wood considered here, which is cut in the L, T plane, it is seen from table 11 that, if breakdown could take place at the edge of the knot it would do so by failure of the fibres in the L direction in compression at about 30° from the section of the plank through the centre of the knot, where $b^2\hat{x}/M = 1.7\pi/b$. Using this approximate value, which is not likely to be altered much by the presence of the straight boundaries if a/b is not too large, it is seen that for $a/b = 0.5$ the maximum value of $b^2\hat{x}/M = 0.85$ and this is well below the maximum at the edge of the plank. It thus seems probable that most knots situated near the centre of the plank will not appreciably alter the conditions of failure.

Referring to table 3a of a previous paper (Green 1942) a similar conclusion is reached when the plank is pierced by a central hole provided that the hole occupies less than a quarter of the section of the plank. For larger holes failure at the hole itself becomes possible.

PART III

THE GENERAL METHOD OF SOLUTION

16. So far attention has been confined to problems of stress distributions in plates of infinite extent which contain a single hole or discontinuity. The presence of other holes may have considerable effect on the distribution of stress and in this part of the paper a method of solution is given for problems of stress distributions in an infinite aeolotropic plate containing any number of circular holes of varying radii. The method of solution is analogous to that used previously (Green 1941) for isotropic materials.

At the centre of each hole a set of rectangular axes is taken parallel to the directions of symmetry of the material. For example, if O_i is the centre

of the r th hole whose radius is a_r , axes are taken at O_r which may be denoted by $O_r x_r, O_r y_r$. Referred to these axes at O_r a complex co-ordinate z_r may be defined as

$$z_r = x_r + iy_r = R_r e^{i\theta_r}, \quad (16.1)$$

where R_r, θ_r are plane polar co-ordinates having their origin at O_r .

If there were no holes in the plate then the stresses produced in it when it is acted on by a system of generalized plane stress may be derived from a stress function χ_0 .* In order to allow for the effect of the holes a fundamental set of stress functions is required for each hole which gives zero stresses at infinity and which gives single-valued expressions for the corresponding stresses and displacements. Suitable functions are those defined in Part II, namely ${}^r\chi_n, {}^r\chi'_n, {}^r\Phi_n, {}^r\Phi'_n$, where the prefix r refers to the r th hole. In these functions u, v will now be denoted by $u_r = z_r + \gamma_1 \bar{z}_r, v_r = z_r + \gamma_2 \bar{z}_r$, and $X_1, X_2, f_n(u), \dots$ become ${}^rX_1, {}^rX_2, {}^rf_n(u), \dots$. If there are m holes the general stress function for any particular problem is then

$$\chi = \chi_0 + \sum_{s=1}^m \sum_{n=0}^{\infty} ({}^sA_n {}^s\chi_n + {}^sC_n {}^s\Phi_n) + \sum_{s=1}^m \sum_{n=1}^{\infty} ({}^sB_n {}^s\chi'_n + {}^sD_n {}^s\Phi'_n), \quad (16.2)$$

where ${}^sA_n, {}^sB_n, {}^sC_n, {}^sD_n$ are constants which are determined by the boundary conditions at the circumference of each hole.

EXPANSIONS ABOUT THE CENTRE OF ANY CIRCLE

17. Before the constants can be evaluated it is necessary to express the stress functions in terms of each set of polar co-ordinates and hence to obtain Fourier expansions for the stresses at the edge of each hole. The stress functions related to the origin O_s produce stresses over the edge of the hole $R_s = a_s$ surrounding the origin O_s which can be expressed in the following forms:

$$\left. \begin{aligned} \widehat{rr}_e &= \sum_{k=0}^{\infty} ({}^na_k^s \cos k\theta_r + {}^nb_k^s \sin k\theta_r), \\ \widehat{r\theta}_e &= \sum_{k=1}^{\infty} ({}^nc_k^s \cos k\theta_r + {}^nd_k^s \sin k\theta_r), \\ \widehat{\theta\theta}_e &= \sum_{k=0}^{\infty} ({}^ne_k^s \cos k\theta_r + {}^nf_k^s \sin k\theta_r), \end{aligned} \right\} \quad (17.1)$$

* This function is not the χ_0 defined in Part II.

$$\begin{aligned} \underline{{}^s\chi'_n} \quad \left. \begin{aligned} \widehat{r\tau}_e &= \sum_{k=0}^{\infty} ({}^ng_k^{s,r} \cos k\theta_r + {}^nh_k^{s,r} \sin k\theta_r), \\ \widehat{r\theta}_e &= \sum_{k=1}^{\infty} ({}^nu_k^{s,r} \cos k\theta_r + {}^nv_k^{s,r} \sin k\theta_r), \\ \widehat{\theta\theta}_e &= \sum_{k=0}^{\infty} ({}^nx_k^{s,r} \cos k\theta_r + {}^ny_k^{s,r} \sin k\theta_r), \end{aligned} \right\} \quad (17.2) \end{aligned}$$

$$\begin{aligned} \underline{{}^s\Phi_n} \quad \left. \begin{aligned} \widehat{r\tau}_e &= \sum_{k=0}^{\infty} ({}^n\alpha_k^{s,r} \cos k\theta_r + {}^n\beta_k^{s,r} \sin k\theta_r), \\ \widehat{r\theta}_e &= \sum_{k=1}^{\infty} ({}^n\gamma_k^{s,r} \cos k\theta_r + {}^n\delta_k^{s,r} \sin k\theta_r), \\ \widehat{\theta\theta}_e &= \sum_{k=0}^{\infty} ({}^n\kappa_k^{s,r} \cos k\theta_r + {}^n\mu_k^{s,r} \sin k\theta_r), \end{aligned} \right\} \quad (17.3) \end{aligned}$$

$$\begin{aligned} \underline{{}^s\Phi'_n} \quad \left. \begin{aligned} \widehat{r\tau}_e &= \sum_{k=0}^{\infty} ({}^n\nu_k^{s,r} \cos k\theta_r + {}^n\tau_k^{s,r} \sin k\theta_r), \\ \widehat{r\theta}_e &= \sum_{k=1}^{\infty} ({}^n\xi_k^{s,r} \cos k\theta_r + {}^n\eta_k^{s,r} \sin k\theta_r), \\ \widehat{\theta\theta}_e &= \sum_{k=0}^{\infty} ({}^n\epsilon_k^{s,r} \cos k\theta_r + {}^n\omega_k^{s,r} \sin k\theta_r). \end{aligned} \right\} \quad (17.4) \end{aligned}$$

The coefficients in these Fourier expansions have not been evaluated for the general case but the values required for the particular problem considered here are given in § 19.

EVALUATION OF THE CONSTANTS

18. The stresses corresponding to χ_0 can be expressed in terms of R_r and θ_r and at the circumference of the circle $R_r = a_r$ they take the forms

$$\widehat{r\tau} = \sum_{k=1}^{\infty} {}^rU_k \sin k\theta_r + \sum_{k=0}^{\infty} {}^rS_k \cos k\theta_r, \quad \widehat{r\theta} = \sum_{k=1}^{\infty} {}^rT_k \sin k\theta_r + \sum_{k=0}^{\infty} {}^rV_k \cos k\theta_r. \quad (18.1)$$

With the help of (17.1)–(17.4) the complete values of $\widehat{r\tau}$ and $\widehat{r\theta}$ at the edge of the hole $R_r = a_r$ may now be found as Fourier series in θ_r . In order that $\widehat{r\tau}$ and $\widehat{r\theta}$ may be zero when $R_r = a_r$ for all values of θ_r the coefficients of the sines and cosines must be zero. These boundary conditions give the following sets of equations:

$${}^rS_0 + {}^rA_0 + \sum_{k=0}^{\infty} \sum_{s=1}^m ({}^sA_k {}^k\alpha_0^{s,r} + {}^sC_k {}^k\alpha_0^{s,r}) + \sum_{k=1}^{\infty} \sum_{s=1}^m ({}^sB_k {}^k\alpha_0^{s,r} + {}^sD_k {}^k\alpha_0^{s,r}) = 0, \quad (18.2)$$

and for $n \geq 1$,

$$\left. \begin{aligned} {}^rS_n + {}^rA_n + {}^rB_n + \sum_{k=0}^{\infty} \sum_{s=1}^m ({}^sA_k {}^k\alpha_n^{s,r} + {}^sC_k {}^k\alpha_n^{s,r}) + \sum_{k=1}^{\infty} \sum_{s=1}^m ({}^sB_k {}^k\alpha_n^{s,r} + {}^sD_k {}^k\alpha_n^{s,r}) &= 0, \\ {}^rT_n + {}^rA_n - {}^rB_n + \sum_{k=0}^{\infty} \sum_{s=1}^m ({}^sA_k {}^k\delta_n^{s,r} + {}^sC_k {}^k\delta_n^{s,r}) + \sum_{k=1}^{\infty} \sum_{s=1}^m ({}^sB_k {}^k\delta_n^{s,r} + {}^sD_k {}^k\delta_n^{s,r}) &= 0. \end{aligned} \right\} \quad (18.3)$$

Also

$${}^rV_0 - {}^rC_0 = 0,$$

and for $n \geq 1$,

$$\left. \begin{aligned} {}^rU_n + {}^rC_n + {}^rD_n + \sum_{k=0}^{\infty} \sum_{s=1}^m ({}^sA_k {}^k\beta_n^{s,r} + {}^sC_k {}^k\beta_n^{s,r}) + \sum_{k=1}^{\infty} \sum_{s=1}^m ({}^sB_k {}^k\beta_n^{s,r} + {}^sD_k {}^k\beta_n^{s,r}) &= 0, \\ {}^rV_n - {}^rC_n + {}^rD_n + \sum_{k=0}^{\infty} \sum_{s=1}^m ({}^sA_k {}^k\gamma_n^{s,r} + {}^sC_k {}^k\gamma_n^{s,r}) + \sum_{k=1}^{\infty} \sum_{s=1}^m ({}^sB_k {}^k\gamma_n^{s,r} + {}^sD_k {}^k\gamma_n^{s,r}) &= 0. \end{aligned} \right\} \quad (18.4)$$

The symbol $\sum_{s=1}^m$ means the sum for all values of s from 1 to m excluding the value $s = r$. The above equations were derived from the boundary conditions at the edge of a typical hole whose centre is O_r . Similar sets of equations may be obtained from the boundary conditions at the edges of all the other holes and so r will take all values from 1 to m .

Since

$$\left. \begin{aligned} {}^k\alpha_1^{s,r} &= {}^k\delta_1^{s,r}, & {}^k\alpha_1^{s,r} &= {}^k\delta_1^{s,r}, & {}^k\alpha_1^{s,r} &= {}^k\delta_1^{s,r}, & {}^k\alpha_1^{s,r} &= {}^k\delta_1^{s,r}, & {}^k\alpha_1^{s,r} &= {}^k\delta_1^{s,r}, \\ {}^k\beta_1^{s,r} &= -{}^k\gamma_1^{s,r}, & {}^k\beta_1^{s,r} &= -{}^k\gamma_1^{s,r}, & {}^k\beta_1^{s,r} &= -{}^k\gamma_1^{s,r}, & {}^k\beta_1^{s,r} &= -{}^k\gamma_1^{s,r}, & {}^k\beta_1^{s,r} &= -{}^k\gamma_1^{s,r}, \end{aligned} \right\} \quad (18.5)$$

$$\text{then} \quad {}^rB_1 = \frac{1}{2}({}^rT_1 - {}^rS_1), \quad {}^rD_1 = -\frac{1}{2}({}^rU_1 + {}^rV_1), \quad (18.6)$$

and a formal solution of equations (18.2)–(18.4) is

$$\left. \begin{aligned} {}^rA_n &= \sum_{p=0}^{\infty} {}^rA_n^{(p)}, & {}^rC_n &= \sum_{p=0}^{\infty} {}^rC_n^{(p)}, \\ {}^rB_n &= \sum_{p=0}^{\infty} {}^rB_n^{(p)}, & {}^rD_n &= \sum_{p=0}^{\infty} {}^rD_n^{(p)}, \end{aligned} \right\} \quad (18.7)$$

where

$${}^rA_0^{(0)} = -{}^rS_0 - \sum_{s=1}^m ({}^sC_0 {}^0\alpha_0^{s,r} + {}^sB_1 {}^1\alpha_0^{s,r} + {}^sD_1 {}^1\alpha_0^{s,r}), \quad (18.8)$$

$$\begin{aligned} {}^rA_n^{(0)} &= -\frac{1}{2}({}^rS_n + {}^rT_n) - \frac{1}{2} \sum_{s=1}^m {}^sC_0 ({}^0\alpha_n^{s,r} + {}^0\delta_n^{s,r}) \\ &\quad - \frac{1}{2} \sum_{s=1}^m \{ {}^sB_1 ({}^1\alpha_n^{s,r} + {}^1\delta_n^{s,r}) + {}^sD_1 ({}^1\alpha_n^{s,r} + {}^1\delta_n^{s,r}) \} \quad (n \geq 1), \end{aligned} \quad (18.9)$$

$$\begin{aligned} {}^rB_n^{(0)} &= \frac{1}{2}({}^rT_n - {}^rS_n) + \frac{1}{2} \sum_{s=1}^m {}^sC_0 ({}^0\delta_n^{s,r} - {}^0\alpha_n^{s,r}) \\ &\quad + \frac{1}{2} \sum_{s=1}^m \{ {}^sB_1 ({}^1\delta_n^{s,r} - {}^1\alpha_n^{s,r}) + {}^sD_1 ({}^1\delta_n^{s,r} - {}^1\alpha_n^{s,r}) \} \quad (n \geq 2), \end{aligned} \quad (18.10)$$

$$\begin{aligned} rC_n^{(0)} = & \frac{1}{2}(rV_n - rU_n) + \frac{1}{2} \sum_{s=1}^m {}^sC_0({}^0\gamma_n^{s,r} - {}^0\beta_n^{s,r}) \\ & + \frac{1}{2} \sum_{s=1}^m \{ {}^sB_1({}^1u_n^{s,r} - {}^1h_n^{s,r}) + {}^sD_1({}^1\xi_n^{s,r} - {}^1\tau_n^{s,r}) \} \quad (n \geq 1), \end{aligned} \quad (18.11)$$

$$\begin{aligned} rD_n^{(0)} = & -\frac{1}{2}(rU_n + rV_n) - \frac{1}{2} \sum_{s=1}^m {}^sC_0({}^0\gamma_n^{s,r} + {}^0\beta_n^{s,r}) \\ & - \frac{1}{2} \sum_{s=1}^m \{ {}^sB_1({}^1h_n^{s,r} + {}^1u_n^{s,r}) + {}^sD_1({}^1\tau_n^{s,r} + {}^1\xi_n^{s,r}) \} \quad (n \geq 2), \end{aligned} \quad (18.12)$$

while for $p \geq 0$

$$\begin{aligned} rA_n^{(p+1)} = & -\frac{1}{2} \sum_{k=0}^{\infty} \sum_{s=1}^m \{ {}^sA_k^{(p)}({}^ka_n^{s,r} + {}^kd_n^{s,r}) + {}^sC_{k+1}^{(p)}({}^{k+1}\alpha_n^{s,r} + {}^{k+1}\delta_n^{s,r}) \\ & + {}^sB_{k+2}^{(p)}({}^{k+2}g_n^{s,r} + {}^{k+2}v_n^{s,r}) + {}^sD_{k+2}^{(p)}({}^{k+2}l_n^{s,r} + {}^{k+2}\eta_n^{s,r}) \} \quad (n \geq 0), \end{aligned} \quad (18.13)$$

$$\begin{aligned} rB_n^{(p+1)} = & \frac{1}{2} \sum_{k=0}^{\infty} \sum_{s=1}^m \{ {}^sA_k^{(p)}({}^kd_n^{s,r} - {}^ka_n^{s,r}) + {}^sC_{k+1}^{(p)}({}^{k+1}\delta_n^{s,r} - {}^{k+1}\alpha_n^{s,r}) \\ & + {}^sB_{k+2}^{(p)}({}^{k+2}v_n^{s,r} - {}^{k+2}g_n^{s,r}) + {}^sD_{k+2}^{(p)}({}^{k+2}\eta_n^{s,r} - {}^{k+2}l_n^{s,r}) \} \quad (n \geq 2), \end{aligned} \quad (18.14)$$

$$\begin{aligned} rC_n^{(p+1)} = & \frac{1}{2} \sum_{k=0}^{\infty} \sum_{s=1}^m \{ {}^sA_k^{(p)}({}^kc_n^{s,r} - {}^kb_n^{s,r}) + {}^sC_{k+1}^{(p)}({}^{k+1}\gamma_n^{s,r} - {}^{k+1}\beta_n^{s,r}) \\ & + {}^sB_{k+2}^{(p)}({}^{k+2}u_n^{s,r} - {}^{k+2}h_n^{s,r}) + {}^sD_{k+2}^{(p)}({}^{k+2}\xi_n^{s,r} - {}^{k+2}\tau_n^{s,r}) \} \quad (n \geq 1), \end{aligned} \quad (18.15)$$

$$\begin{aligned} rD_n^{(p+1)} = & -\frac{1}{2} \sum_{k=0}^{\infty} \sum_{s=1}^m \{ {}^sA_k^{(p)}({}^kb_n^{s,r} + {}^kc_n^{s,r}) + {}^sC_{k+1}^{(p)}({}^{k+1}\beta_n^{s,r} + {}^{k+1}\gamma_n^{s,r}) \\ & + {}^sB_{k+2}^{(p)}({}^{k+2}h_n^{s,r} + {}^{k+2}u_n^{s,r}) + {}^sD_{k+2}^{(p)}({}^{k+2}\tau_n^{s,r} + {}^{k+2}\xi_n^{s,r}) \} \quad (n \geq 2). \end{aligned} \quad (18.16)$$

Equation (18.13) includes the case $n = 0$ if the constants ${}^kd_0^{s,r}, \dots$ are defined to be

$${}^kd_0^{s,r} = {}^ka_0^{s,r}, \quad {}^k\delta_0^{s,r} = {}^k\alpha_0^{s,r}, \quad {}^kv_0^{s,r} = {}^kg_0^{s,r}, \quad {}^k\eta_0^{s,r} = {}^kl_0^{s,r}. \quad (18.17)$$

No attempt will be made here to establish the convergence of the above formal solution.

TWO EQUAL CIRCULAR HOLES

19. The analysis developed in §§ 16–18 will now be applied to the particular case of two equal holes situated so that their line of centres is either along or perpendicular to the grain. Axes $O_1x_1, O_1y_1; O_2x_2, O_2y_2$, are taken at the centres of the circles as shown in figure 13 so that if $O_1O_2 = b$,

$$z_1 + z_2 + b = 0. \quad (19.1)$$

Attention will be confined to stress systems which are symmetrical about the line of centres O_1O_2 and the perpendicular bisector of O_1O_2 so that the stress function (16.2) reduces to

$$\chi = \chi_0 + \sum_{n=0}^{\infty} {}^1A_n({}^1\chi_n + {}^2\chi_n) + \sum_{n=2}^{\infty} {}^1B_n({}^1\chi'_n + {}^2\chi'_n). \quad (19.2)$$

In order to complete the solution of any given problem it is necessary to obtain the values of the coefficients given in § 17 which arise from the Fourier expansions of the values of the stresses over the hole $R_1 = a$ due to stress functions which are related to the other hole. For this purpose Fourier expansions of expressions of the type

$$\frac{\lambda^n(z_2 + \gamma\bar{z}_2)^n \left[1 - \left\{ 1 - \frac{4a^2\gamma}{(z_2 + \gamma\bar{z}_2)^2} \right\}^{\frac{1}{2}} \right]^{n+1}}{a^n \left\{ 1 - \frac{4a^2\gamma}{(z_2 + \gamma\bar{z}_2)^2} \right\}^{\frac{1}{2}}} \quad (19.3)$$

are required in terms of θ_1 when $R_1 = a$.

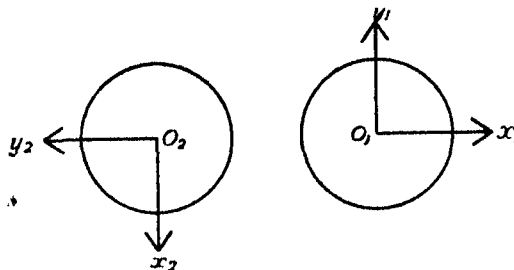


FIGURE 13

For convenience the following notation is introduced:

$$\lambda = a\{b^2(1+\gamma)^2 - 4a^2\gamma\}^{-\frac{1}{2}}, \quad \mu = b(1+\gamma)\{b^2(1+\gamma)^2 - 4a^2\gamma\}^{-\frac{1}{2}}, \quad (19.4)$$

$$w = \lambda(e^{i\theta_1} + \gamma e^{-i\theta_1}),$$

and a suffix 1 or 2 is added to λ, μ, w to correspond with γ equal to γ_1 or γ_2 respectively.

When $z_1 = ae^{i\theta_1}$, the expression (19.3) can, with the help of (19.1) and (19.4), be written as

$$\Phi(\mu, w) = \frac{(\mu + w)^n \left[1 - \left\{ \frac{1 + 2\mu w + w^2}{\mu^2 + 2\mu w + w^2} \right\}^{\frac{1}{2}} \right]^{n+1}}{\left\{ \frac{1 + 2\mu w + w^2}{\mu^2 + 2\mu w + w^2} \right\}^{\frac{1}{2}}}, \quad (19.5)$$

if a factor $(-)^n$ is omitted. It can be verified that $\Phi(\mu, w)$ satisfies the partial differential equation

$$(\mu^2 - 1) \frac{\partial^2 \Phi}{\partial \mu^2} - 2n\mu \frac{\partial \Phi}{\partial \mu} + n(n+1)\Phi - w \frac{\partial^2}{\partial w^2} (w\Phi) = 0, \quad (19.6)$$

and if the substitution

$$\Phi(\mu, w) = (\mu^2 - 1)^{n+1} \Psi(\mu, w) \quad (19.7)$$

is made then $\Psi(\mu, w)$ satisfies the equation

$$(\mu^2 - 1) \frac{\partial^2 \Psi}{\partial \mu^2} + 2(n+2)\mu \frac{\partial \Psi}{\partial \mu} + (n+1)(n+2)\Psi - w \frac{\partial^2}{\partial w^2} (w\Psi) = 0. \quad (19.8)$$

For the values of μ, w given in (19.4) $\Psi(\mu, w)$ is a regular function of w which can be expanded in a Taylor series in the form

$$\Psi(\mu, w) = \sum_{r=0}^{\infty} w^r G_r(\mu). \quad (19.9)$$

This series may be differentiated any number of times with respect to w and it may also be shown that it may be differentiated at least twice with respect to μ in the possible range of values of μ, w . Thus the series (19.9) will satisfy equation (19.8) provided that

$$(1 - \mu^2) \frac{\partial^2 G_r}{\partial \mu^2} - 2\mu(n+2) \frac{\partial G_r}{\partial \mu} + (r-n-1)(r+n+2) G_r = 0. \quad (19.10)$$

The relevant solution of this equation is

$$G_r(\mu) = A F(r+n+2, n+1-r; n+2; \delta) \quad (19.11)$$

where $F(r+n+2, n+1-r; n+2; \delta)$ is the ordinary hyper-geometric function, A is a numerical factor which depends on r and n and $\delta = \frac{1}{2}(1 - \mu)$. The factor A can be determined by considering the expansion of $\Psi(\mu, w)$ when $\mu = 1$. It is found that

$$G_r(\mu) = \frac{(-)^r (r+n+1)!}{r! (n+1)! 2^{n+1}} F(r+n+2, n+1-r; n+2; \delta), \quad (19.12)$$

and this may also be expressed in the form

$$G_r(\mu) = \frac{(-)^r (r+n+1)!}{r! (n+1)! (\mu+1)^{n+1}} F(-r, r+1; n+2; \delta). \quad (19.13)$$

The above expansion may also be obtained by using Lagrange's expansion formula (see Hobson 1931) which gives $G_r(\mu)$ in terms of the associated Legendre function $P_r^{n+1}(\mu)$.

When $R_1 = a$ the required expansion of (19.3) now takes the form

$$\frac{(z_2 + \gamma \bar{z}_2)^n \left[1 - \left\{ 1 - \frac{4a^2 \gamma}{(z_2 + \gamma \bar{z}_2)^2} \right\}^{\frac{1}{2}} \right]^{n+1}}{a^n \left\{ 1 - \frac{4a^2 \gamma}{(z_2 + \gamma \bar{z}_2)^2} \right\}^{\frac{1}{2}}} \\ = \frac{(-)^n (\mu - 1)^{n+1}}{\lambda^n (n+1)!} \sum_{r=0}^{\infty} \frac{(-)^r (r+n+1)! w^r}{r!} F(-r, r+1; n+2; \delta). \quad (19.14)$$

Since the applications are confined to those stress systems which are symmetrical about the line of centres and its perpendicular bisector, not all of the coefficients in (17.1)–(17.4) are required and only the relevant ones will be given here. The coefficients for unsymmetrical systems may easily be obtained by a similar method if needed. Consider the stress functions ${}^2\chi_n$. With the help of (3.2), (3.3) and (19.4) it will be found that the value of the normal stress $\hat{r}\hat{r}$ over the edge of the hole whose centre is O_1 , due to ${}^2\chi_n$, is given by

$$\hat{r}\hat{r} = \frac{(-)^{n+1}}{(\gamma_1 - \gamma_2) 2^{n+1} (n+1)!} \sum_{r=0}^{\infty} \frac{(-)^r (r+n+1)!}{r!} \\ \times \left\{ \frac{\gamma_2 (\mu_1 - 1)^{n+1} \lambda_1^{r-n}}{\gamma_1^{n+1}} (e^{i\theta_1} + \gamma_1 e^{-i\theta_1})^r \{ (e^{i\theta_1} + \gamma_1 e^{-i\theta_1})^2 - 4\gamma_1 \} F(-r, r+1; n+2; \delta_1) \right. \\ \left. - \frac{\gamma_1 (\mu_2 - 1)^{n+1} \lambda_2^{r-n}}{\gamma_2^{n+1}} (e^{i\theta_1} + \gamma_2 e^{-i\theta_1})^r \{ (e^{i\theta_1} + \gamma_2 e^{-i\theta_1})^2 - 4\gamma_2 \} F(-r, r+1; n+2; \delta_2) \right\}. \quad (19.15)$$

Each term in θ_1 in (19.15) must now be expanded and, considering only the real part, the expression can be rearranged as a Fourier cosine series whose coefficients are

$${}^n a_{2k}^{2,1} = \frac{(-)^n}{(\gamma_1 - \gamma_2) 2^n (n+1)!} \sum_{r=k-1}^{\infty} \frac{(r+1-2k^2)(2r+n+1)!}{(r+1+k)!(r+1-k)!} \\ \times \{ \gamma_2 (\mu_1 - 1)^{n+1} (1 + \gamma_1^{2k}) \lambda_1^{2r-n} \gamma_1^{r-n-k} F(-2r, 2r+1; n+2; \delta_1) \\ - \gamma_1 (\mu_2 - 1)^{n+1} (1 + \gamma_2^{2k}) \lambda_2^{2r-n} \gamma_2^{r-n-k} F(-2r, 2r+1; n+2; \delta_2) \} \quad (k \geq 0), \quad (19.16)$$

$${}^n a_{2k+1}^{2,1} = \frac{(-)^{n+1}}{(\gamma_1 - \gamma_2) 2^n (n+1)!} \sum_{r=k-1}^{\infty} \frac{(r+1-2k-2k^2)(2r+n+2)!}{(r+2+k)!(r+1-k)!} \\ \times \{ \gamma_2 (\mu_1 - 1)^{n+1} (1 + \gamma_1^{2k+1}) \lambda_1^{2r+1-n} \gamma_1^{r-n-k} F(-2r-1, 2r+2; n+2; \delta_1) \\ - \gamma_1 (\mu_2 - 1)^{n+1} (1 + \gamma_2^{2k+1}) \lambda_2^{2r+1-n} \gamma_2^{r-n-k} F(-2r-1, 2r+2; n+2; \delta_2) \} \quad (k \geq 0). \quad (19.17)$$

The remaining coefficients in (17.1) and (17.2) may be found by a similar process. The details of the work will be omitted and only the final results will be quoted here. Thus, for $k \geq 0$, $n \geq 0$,

$$\begin{aligned} {}^n d_{2k}^{2,1} = & \frac{(-)^n}{(\gamma_1 - \gamma_2) 2^n (n+1)!} \sum_{r=k-1}^{\infty} \frac{k(2r+1)(2r+n+1)!}{(r+1+k)! (r+1-k)!} \\ & \times \{ \gamma_2 (\mu_1 - 1)^{n+1} (1 + \gamma_1^{2k}) \lambda_1^{2r-n} \gamma_1^{r-n-k} F(-2r, 2r+1; n+2; \delta_1) \\ & - \gamma_1 (\mu_2 - 1)^{n+1} (1 + \gamma_2^{2k}) \lambda_2^{2r-n} \gamma_2^{r-n-k} F(-2r, 2r+1; n+2; \delta_2) \}, \end{aligned} \quad (19.18)$$

$$\begin{aligned} {}^n d_{2k+1}^{2,1} = & \frac{(-)^{n+1}}{(\gamma_1 - \gamma_2) 2^n (n+1)!} \sum_{r=k-1}^{\infty} \frac{(2k+1)(r+1)(2r+n+2)!}{(r+2+k)! (r+1-k)!} \\ & \times \{ \gamma_2 (\mu_1 - 1)^{n+1} (1 + \gamma_1^{2k+1}) \lambda_1^{2r+1-n} \gamma_1^{r-n-k} F(-2r-1, 2r+2; n+2; \delta_1) \\ & - \gamma_1 (\mu_2 - 1)^{n+1} (1 + \gamma_2^{2k+1}) \lambda_2^{2r+1-n} \gamma_2^{r-n-k} F(-2r-1, 2r+2; n+2; \delta_2) \}, \end{aligned} \quad (19.19)$$

$$\begin{aligned} {}^n e_{2k}^{2,1} = & \frac{(-)^n}{(\gamma_1 - \gamma_2) 2^n (n+1)!} \sum_{r=k-1}^{\infty} \frac{(r+1)(2r+1)(2r+n+1)!}{(r+1+k)! (r+1-k)!} \\ & \times \{ \gamma_2 (\mu_1 - 1)^{n+1} (1 + \gamma_1^{2k}) \lambda_1^{2r-n} \gamma_1^{r-n-k} F(-2r, 2r+1; n+2; \delta_1) \\ & - \gamma_1 (\mu_2 - 1)^{n+1} (1 + \gamma_2^{2k}) \lambda_2^{2r-n} \gamma_2^{r-n-k} F(-2r, 2r+1; n+2; \delta_2) \}, \end{aligned} \quad (19.20)$$

$$\begin{aligned} {}^n e_{2k+1}^{2,1} = & \frac{(-)^{n+1}}{(\gamma_1 - \gamma_2) 2^n (n+1)!} \sum_{r=k-1}^{\infty} \frac{(r+1)(2r+3)(2r+n+2)!}{(r+2+k)! (r+1-k)!} \\ & \times \{ \gamma_2 (\mu_1 - 1)^{n+1} (1 + \gamma_1^{2k+1}) \lambda_1^{2r+1-n} \gamma_1^{r-n-k} F(-2r-1, 2r+2; n+2; \delta_1) \\ & - \gamma_1 (\mu_2 - 1)^{n+1} (1 + \gamma_2^{2k+1}) \lambda_2^{2r+1-n} \gamma_2^{r-n-k} F(-2r-1, 2r+2; n+2; \delta_2) \}, \end{aligned} \quad (19.21)$$

and for $k \geq 0$, $n \geq 2$

$$\begin{aligned} {}^n g_{2k}^{2,1} = & \frac{(-)^n}{(\gamma_1 - \gamma_2) 2^{n-2} (n-1)!} \sum_{r=k-1}^{\infty} \frac{(r+1-2k^2)(2r+n-1)!}{(r+1+k)! (r+1-k)!} \\ & \times \{ (\mu_1 - 1)^{n-1} (1 + \gamma_1^{2k}) \lambda_1^{2r-n-2} \gamma_1^{r-n-k-2} F(-2r, 2r+1; n; \delta_1) \\ & - (\mu_2 - 1)^{n-1} (1 + \gamma_2^{2k}) \lambda_2^{2r-n-2} \gamma_2^{r-n-k-2} F(-2r, 2r+1; n; \delta_2) \}, \end{aligned} \quad (19.22)$$

$$\begin{aligned} {}^n g_{2k+1}^{2,1} = & \frac{(-)^{n+1}}{(\gamma_1 - \gamma_2) 2^{n-2} (n-1)!} \sum_{r=k-1}^{\infty} \frac{(r+1-2k-2k^2)(2r+n)!}{(r+2+k)! (r+1-k)!} \\ & \times \{ (\mu_1 - 1)^{n-1} (1 + \gamma_1^{2k+1}) \lambda_1^{2r-n-1} \gamma_1^{r-n-k-2} F(-2r-1, 2r+2; n; \delta_1) \\ & - (\mu_2 - 1)^{n-1} (1 + \gamma_2^{2k+1}) \lambda_2^{2r-n-1} \gamma_2^{r-n-k-2} F(-2r-1, 2r+2; n; \delta_2) \}, \end{aligned} \quad (19.23)$$

$$\begin{aligned} {}^n v_{2k}^{2,1} = & \frac{(-)^n}{(\gamma_1 - \gamma_2) 2^{n-2} (n-1)!} \sum_{r=k-1}^{\infty} \frac{k(2r+1)(2r+n-1)!}{(r+1+k)! (r+1-k)!} \\ & \times \{ (\mu_1 - 1)^{n-1} (1 + \gamma_1^{2k}) \lambda_1^{2r-n-2} \gamma_1^{r-n-k-2} F(-2r, 2r+1; n; \delta_1) \\ & - (\mu_2 - 1)^{n-1} (1 + \gamma_2^{2k}) \lambda_2^{2r-n-2} \gamma_2^{r-n-k-2} F(-2r, 2r+1; n; \delta_2) \}, \end{aligned} \quad (19.24)$$

$$\begin{aligned}
{}^nv_{2k+1}^{2,1} = & \frac{(-)^{n+1}}{(\gamma_1 - \gamma_2) 2^{n-2} (n-1)!} \sum_{r=k-1}^{\infty} \frac{(2k+1)(r+1)(2r+n)!}{(r+2+k)!(r+1-k)!} \\
& \times \{(\mu_1 - 1)^{n-1} (1 + \gamma_1^{2k+1}) \lambda_1^{2r-1-n} \gamma_1^{r-n-k-2} F(-2r-1, 2r+2; n; \delta_1) \\
& - (\mu_2 - 1)^{n-1} (1 + \gamma_2^{2k+1}) \lambda_2^{2r-1-n} \gamma_2^{r-n-k-2} F(-2r-1, 2r+2; n; \delta_2)\}, \quad (19.25)
\end{aligned}$$

$$\begin{aligned}
{}^nx_{2k}^{2,1} = & \frac{(-)^n}{(\gamma_1 - \gamma_2) 2^{n-2} (n-1)!} \sum_{r=k-1}^{\infty} \frac{(r+1)(2r+1)(2r+n-1)!}{(r+1+k)!(r+1-k)!} \\
& \times \{(\mu_1 - 1)^{n-1} (1 + \gamma_1^{2k}) \lambda_1^{2r-n-2} \gamma_1^{r-n-k-2} F(-2r, 2r+1; n; \delta_1) \\
& - (\mu_2 - 1)^{n-1} (1 + \gamma_2^{2k}) \lambda_2^{2r-n-2} \gamma_2^{r-n-k-2} F(-2r, 2r+1; n; \delta_2)\}, \quad (19.26)
\end{aligned}$$

$$\begin{aligned}
{}^nx_{2k+1}^{2,1} = & \frac{(-)^{n+1}}{(\gamma_1 - \gamma_2) 2^{n-2} (n-1)!} \sum_{r=k-1}^{\infty} \frac{(r+1)(2r+3)(2r+n)!}{(r+2+k)!(r+1-k)!} \\
& \times \{(\mu_1 - 1)^{n-1} (1 + \gamma_1^{2k+1}) \lambda_1^{2r-1-n} \gamma_1^{r-n-k-2} F(-2r-1, 2r+2; n; \delta_1) \\
& - (\mu_2 - 1)^{n-1} (1 + \gamma_2^{2k+1}) \lambda_2^{2r-1-n} \gamma_2^{r-n-k-2} F(-2r-1, 2r+2; n; \delta_2)\}. \quad (19.27)
\end{aligned}$$

The remaining coefficients in (17.1) and (17.2) are zero for two holes whose line of centres is along or perpendicular to the grain. The set of coefficients ${}^na_k^{1,2}, \dots$ is equal to the set ${}^na_k^{2,1}, \dots$ given above. The values of ${}^nd_0^{2,1}$ and ${}^nv_0^{2,1}$ are not given by (19.18) and (19.24) but are to be obtained from (18.17).

THE TENSION PROBLEM

20. If the plate is under a uniform tension T parallel to the line of centres, then

$${}^1S_0 = {}^2S_0 = {}^1S_2 = {}^2S_2 = -{}^1T_2 = -{}^2T_2 = \frac{1}{2}T, \quad (20.1)$$

and

$${}^1A_0^{(0)} = -\frac{1}{2}T, \quad {}^1A_n^{(0)} = 0 \quad (n \geq 1), \quad {}^1B_2^{(0)} = -\frac{1}{2}T, \quad {}^1B_n^{(0)} = 0 \quad (n \geq 3). \quad (20.2)$$

Calculations for any given material and for particular arrangements of the hole are long. In order to get some idea of the effect of one hole on the stress distribution round a neighbouring hole calculations have been made, with the help of §§ 17–19, for spruce when the grain is perpendicular to the line of centres and for the case when the distance between the centres of the holes is four times the radius of either hole, i.e. $b = 4a$. The tabulation of the various constants which are needed in this work would take up considerable space, and they are of no use for any other material or arrangement of holes, so only the final values of the edge stresses round either hole are given in table 12 where $\theta = 0$ is parallel to the line of centres. Values of $\partial\theta$ for a single hole in spruce are included for comparison. It will be seen that

the presence of the second hole tends to reduce the magnitudes of the stress round most of the hole, the reduction being most marked in the compressive stress at points of the two holes which are nearest together.

TABLE 12. VALUES OF STRESSES ON THE EDGE OF A HOLE. TENSION
PARALLEL TO LINE OF CENTRES AND PERPENDICULAR TO GRAIN

θ°	\widehat{xx}/T	\widehat{yy}/T	\widehat{xy}/T	$\widehat{\theta\theta}/T$	$\widehat{\theta\theta}/T^*$
0	0	-4.86	0	-4.86	-5.14
5	-0.030	-3.89	0.340	-3.92	—
10	-0.062	-2.01	0.354	-2.07	-2.22
15	-0.031	-0.433	0.116	-0.464	—
20	0.072	0.546	-0.199	0.618	0.581
30	0.420	1.26	-0.727	1.68	1.70
40	0.835	1.19	-0.995	2.02	2.07
50	1.23	0.864	-1.03	2.09	2.15
60	1.55	0.515	-0.892	2.06	2.13
70	1.77	0.235	-0.646	2.01	2.09
80	1.91	0.059	-0.337	1.97	2.06
90	1.94	0	0	1.94	2.04
100	1.87	0.058	0.330	1.93	2.06
110	1.71	0.227	0.624	1.94	2.09
120	1.46	0.485	0.840	1.94	2.13
130	1.13	0.793	0.945	1.92	2.15
140	0.744	1.06	0.886	1.80	2.07
150	0.363	1.09	0.628	1.45	1.70
160	0.057	0.429	0.156	0.486	0.581
165	-0.031	-0.429	-0.115	-0.460	—
170	-0.056	-1.79	-0.316	-1.85	-2.22
175	-0.026	-3.40	-0.298	-3.43	—
180	0	-4.23	0	-4.23	-5.14

* Single hole.

21. When a tension is applied at right angles to the line of centres

$${}^1S_0 = {}^2S_0 = {}^1T_2 = {}^2T_2 = -{}^1S_2 = -{}^2S_2 = \frac{1}{2}T, \quad (21.1)$$

and

$${}^1A_0^{(0)} = -\frac{1}{2}T, \quad {}^1A_n^{(0)} = 0 \quad (n \geq 1), \quad {}^1B_2^{(0)} = \frac{1}{2}T, \quad {}^1B_n^{(0)} = 0 \quad (n \geq 3). \quad (21.2)$$

Calculations are again limited to a spruce plank whose grain is perpendicular to the line of centres and for the case $b = 4a$. The stresses at the edge of either hole are shown in table 13 where $\theta = 0$ is parallel to the line of centres and the values for a single hole are included for comparison. In this

case there is a general increase in the values of the tensional edge stresses due to the presence of the second hole, the increase being greatest at the points of the holes which are nearest together although these are not the points at which there is the greatest percentage increase. The compressive stress is slightly less for two holes than for one.

TABLE 13. VALUES OF STRESSES ON THE EDGE OF A HOLE. TENSION PERPENDICULAR TO LINE OF CENTRES AND PARALLEL TO GRAIN

θ°	\widehat{xx}/T	\widehat{yy}/T	\widehat{xy}/T	$\widehat{\theta\theta}/T$	$\widehat{\theta\theta}/T^*$
0	0	6.52	0	6.52	6.37
5	0.0434	5.67	-0.496	5.71	5.57
10	0.124	3.98	-0.701	4.10	3.99
15	0.180	2.50	-0.670	2.68	2.61
20	0.199	1.50	-0.546	1.70	1.65
30	0.161	0.482	-0.278	0.643	0.615
40	0.0793	0.113	-0.0945	0.192	0.173
50	-0.009	-0.006	0.007	-0.015	-0.028
60	-0.084	-0.028	0.0485	-0.112	-0.123
70	-0.140	-0.0185	0.0508	-0.158	-0.169
80	-0.174	-0.005	0.0306	-0.179	-0.189
90	-0.181	0	0	-0.181	-0.195
100	-0.166	-0.005	-0.029	-0.171	-0.189
110	-0.129	-0.017	-0.047	-0.146	-0.169
120	-0.070	-0.023	-0.040	-0.093	-0.123
130	0.008	0.006	0.007	0.014	-0.028
140	0.095	0.136	0.114	0.231	0.173
150	0.174	0.523	0.302	0.697	0.615
160	0.207	1.56	0.569	1.77	1.65
165	0.184	2.57	0.688	2.75	2.61
170	0.125	4.02	0.710	4.15	3.99
175	0.044	5.72	0.500	5.76	5.57
180	0	6.57	0	6.57	6.37

* Single hole.

In general it may be concluded that the stress distribution round a hole in a spruce tension member is not greatly affected by the presence of another hole provided that the line of centres of the holes is perpendicular to the grain and to the tension, and provided that the shortest distance between the boundaries of the holes is not smaller than the diameter of either hole. When the tension is parallel to the line of centres the effect of the second hole is more marked but the magnitudes of the stresses are mostly less than for a single hole.

REFERENCES

- Coker, E. G. & Filon, L. N. G. 1931 *A treatise on photo-elasticity*. Camb. Univ. Press.
 Donnell, L. H. 1941 *Theodore Von Karman Anniversary volume*, p. 203.
 Green, A. E. 1940 *Proc. Roy. Soc. A*, **176**, 121.
 Green, A. E. 1941 *Proc. Camb. Phil. Soc.* **37**, 29.
 Green, A. E. 1942 *Proc. Roy. Soc. A*, **180**, 173.
 Green, A. E. 1943 *Phil. Mag.* **34**, 416.
 Green, A. E. 1945 *Proc. Roy. Soc. A*, **184**, 231.
 Green, A. E. & Taylor, G. I. 1939 *Proc. Roy. Soc. A*, **173**, 162.
 Green, A. E. & Taylor, G. I. 1945 *Proc. Roy. Soc. A*, **184**, 181.
 Gurney, C. 1938 *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1834.
 Hobson, E. W. 1931 *The theory of spherical and ellipsoidal harmonics*. Camb. Univ. Press.
 Howland, R. C. J. & Stevenson, A. C. 1933 *Phil. Trans. A*, **232**, 155.
 Muschelisvili, N. 1933 *Z. angew. Math. Mech.* **13**, 264.
 Suyehiro, K. 1914 *J. Soc. Mech. Engrs Japan*.
 Stevenson 1945 *Proc. Roy. Soc. A*, **184**, 218.

The phosphorescence of various solids

By J. T. RANDALL

Royal Society Warren Research Fellow

AND M. H. F. WILKINS

Physics Department, University of Birmingham

(Communicated by M. L. E. Oliphant, F.R.S.—Received 18 December 1939)

[Plate 8]

INTRODUCTION

Ultra-violet radiation absorbed by a solid may be dissipated either as heat or as luminescence. If the electrons involved in the absorption transitions return to the ground state in a time not greater than 10^{-6} sec., the resulting emission is described as fluorescence. Delayed return is generally referred to as phosphorescence, and it is the chief object of this paper to describe some measurements of the decay law for various solids, together with some associated measurements of photoconductivity. A preliminary account of this work has already been published (Randall and Wilkins 1939).

The phenomenon of phosphorescence can be described by reference to figure 1. Suppose we have a beam of ultra-violet radiation the intensity of which varies with time as shown in figure 1a. If this beam is allowed to fall



FIGURE 1a

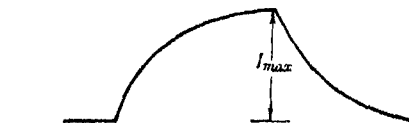


FIGURE 1b

on a solid and excite luminescence, the intensity of luminescence will vary as shown in figure 1b, rising to an equilibrium value I_{\max} , according to definite laws, and decaying after the exciting radiation has been cut off, also according to definite laws (Randall 1939; de Groot 1939a). The law of decay is in general a complex one; the two simplest assumptions concerning the mechanism of decay give an exponential and a hyperbolic law. Further reference to this is made in the discussion. The time constants for

the rise and decay for any chosen solid appear to be the same, or approximately so, in all cases examined so far. The decay periods vary over a wide range. The decay in platinocyanides and a number of tungstates appears from approximate measurements to be complete in a time of 10^{-5} sec. In zinc and zinc-cadmium sulphides the decay may last for minutes, and in alkaline earth sulphides periods of many hours or days have been observed.

Since the experiments of Gudden and Pohl (1921) it has been known that there may exist some direct connexion between luminescence and photoconductivity, and modern ideas of energy states in solids have shown in what way such a connexion could arise. Few attempts have so far been made to probe this connexion further. The optical energy levels in solids are complex, and it is necessary for a complete understanding of the mechanism involved in luminescence processes to approach the problem from as many directions as possible. In addition to the phosphorescence measurements already mentioned, we have also examined the photoconductivity of our specimens. The results show some complexity; in the discussion which follows the presentation of results, we examine a number of the more probable explanations.

EXPERIMENTAL TECHNIQUE

The fundamental principle of the apparatus is in no way original. The metal disk *D* (figures 2, 3) 10 in. in diameter rotated on ball bearings, the axis of rotation being horizontal. The disk was driven by a d.c. motor and endless web-belt attached to the pulley *P*. The speed was adjusted by a resistance in series with the motor, and the speed could be kept constant by using a neon lamp running on the a.c. mains as a stroboscope. Speeds of 2000 r.p.m. were used for the uranyl salts, and speeds of 125 and 250 r.p.m. for the impurity phosphors. It was important to adjust the speed of the disk so that the phosphorescence decayed to a negligible value after one revolution.

In order to obtain as much light as possible the powdered specimens were applied in annulus form round the circumference of the disk. We found it convenient to apply the powder to a cellophane disk which was fixed to the metal disk by adhesive. (For speeds up to 2000 r.p.m., vacuum tap grease is useful as it allows the specimen sheet to be removed intact from the wheel. Tap grease was also used for holding the powder to the cellophane.)

The exciting source was an 80 watt Osira mercury lamp (with outer glass bulb removed) working on d.c. A single-stage quartz monochromator was used to transmit only a wide group of lines about 2560 Å. The light from

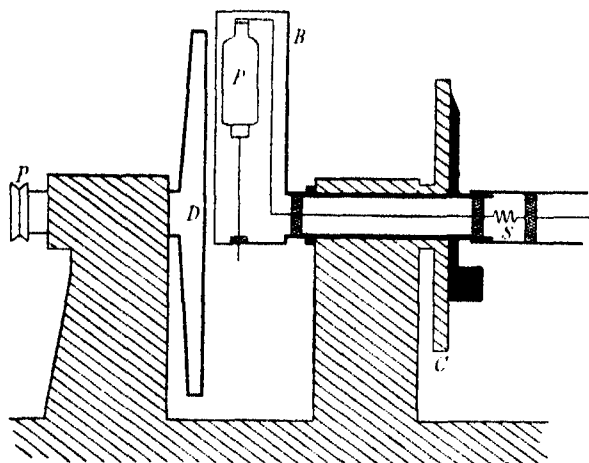


FIGURE 2. Side view of disk phosphoroscope, showing rotating disk *D*, with photocell *P* and divided circle *C*.

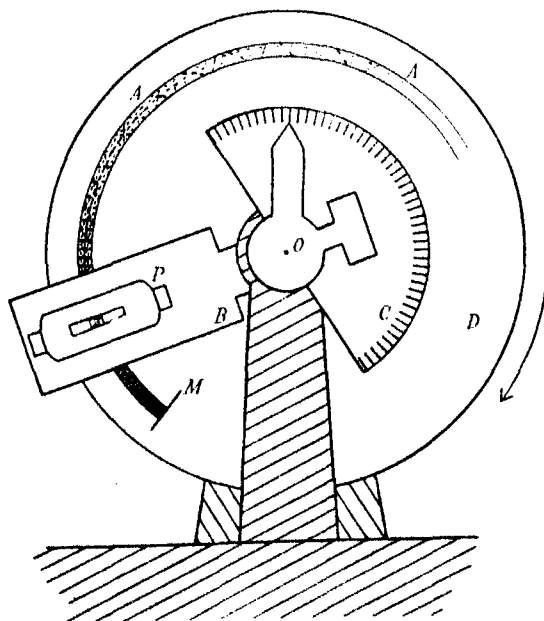


FIGURE 3. Front view of disk phosphoroscope showing photocell *P* in box *B* arranged for reading phosphorescence intensity at any point of annulus *AA*. Luminescence is excited at *M* by ultra-violet light.

the monochromator was focused in a small patch at *M* (figure 3). Stray ultra-violet light was cut off by diaphragms. An attempt was made to use an α -particle source of about 200 millicuries of radon but insufficient phosphorescence was excited.

The intensity of phosphorescence was measured by means of a vacuum photocell *P* (type K.M.V. 6, G.E.C. potassium cell) and amplifier of the electrometer triode type (Winch and Machin 1935; Warren 1935). The advantage of the vacuum photocell and triode method lies in its sensitivity, steadiness, and linear response to light intensity. The deflexion of a galvanometer measured the light intensity directly. The photocell was mounted in an earth-shielded box *B* on an arm rotating on an axis *O* in line with the axis of rotation of the phosphorescence disk. A divided circle *C* allowed the angular position of the cell to be read to within 0.1° . Lead-wires to the amplifier were kept short and were shielded, the rotating lead connected to the stationary lead through a fine spiral of wire *S*, figure 2. Wratten 2a filters were used to cover the slit on the photocell box so that stray ultra-violet light entering was negligible. The dimensions of the slit on the photocell box were 3×35 mm. With the box 2 mm. from the phosphoroscope disk the intensity recorded for a given setting of the box was a measure of the average intensity over 2° round the phosphorescent annulus.

Observation of the decay started at approximately 14° after the exciting light was cut off and thus the first part of the decay curve could not be observed. This was inevitable as the photocell box would cut off the exciting light if the photocell was moved closer to the beginning of the phosphorescence tail. Readings of intensity reproduced to within 1% for the larger scale deflexions. A photograph of the phosphorescence decay of cadmium chlorophosphate on the rotating disk is shown in figure 4, plate 1 (the photocell has been removed to show the complete decay).

Oscillograph phosphoroscope

In addition to the apparatus just described we have also developed an oscillograph phosphoroscope, essentially similar to those developed independently by Schleede and Bartels (1938), and de Groot (1939*a, b*). In this type of apparatus a constant source of ultra-violet radiation is regularly interrupted by means of a sector disk and allowed to fall on the specimen to be investigated. The luminescent light is focused on the primary cathode of a multistage electron multiplier. The potential developed across the load resistance in the multiplier circuit is amplified by means of a linear amplifier and applied to an oscillograph. The rise and decay curves are seen on the screen and may be photographed for measurement purposes. One ad-



FIGURE 4. Photograph of rotating disk phosphoroscope showing decay of cadmium chlorophosphate.



FIGURE 5. Rise and decay of ammonium uranyl phosphate. Above 90° K; below 300° K.



FIGURE 6. Decay of cadmium chlorophosphate. Above 90° K; below 300° K.



FIGURE 7. Decay of zinc beryllium silicate. Above 90° K; below 300° K.



FIGURE 8. Decay of zinc orthosilicate (Mn). Above 90° K; centre 300° K; below 520° K.

vantage of this type of phosphoroscope is that major effects of temperature, exciting intensity etc. can be observed rapidly. Examples of oscillograph records of various materials are shown in figures 5-8, plate 1. Our use of the oscillograph phosphoroscope has so far been chiefly qualitative.

THE PHOSPHORESCENCE OF URANYL SALTS

The fluorescence spectra of the uranyl salts consist of a series of narrow bands. This in itself indicates that emission is the result of electron transitions from excitation states in the crystal. Solutions of the salts fluoresce strongly. This and other evidence indicates that the co-ordination group of the uranyl ion is the seat of luminescence, whether the ion is in a solvent or part of the pure crystal. That the luminescent process is not confined to special centres is indicated by the observation that the fluorescence intensity does not tend to saturate for high intensities of exciting light as it does for impurity phosphors (de Groot 1939c). It is very reasonable to expect for uranyl salts exponential phosphorescence decays independent of exciting intensity and no photoconductivity. This is found to be the case.

Wawilow and Lewschin (1928), using visual photometry, showed that the uranyl salts give exponential decays. They also showed that the results of Nichols and Howes (1919) could be interpreted to show exponential decays.

Figure 9, curve *e*, shows our results for ammonium uranyl phosphate. The decay constant α in the formula $I_t = I_0 e^{-\alpha t}$ is 703 sec.⁻¹. The log of the intensity is plotted against time. In figure 10 the curves show that the decay of the phosphate and nitrate is unaffected by variation of the intensity of the exciting light. The decay constant for uranyl nitrate is 1390 sec.⁻¹. Wawilow and Lewschin give the value 1800 sec.⁻¹. This discrepancy possibly arises from difficulties of accurate visual photometry as used by Wawilow and Lewschin.

THE PHOSPHORESCENCE OF COMPOUNDS WITH MANGANESE IMPURITY

The phosphorescence of a number of solids containing manganese impurity has been examined and a list of these is given below.

The first four of the compounds in this table give almost exponential decay curves, the semi-log plots being shown in figure 9*a, b, c, d*. It was at first thought from preliminary examination that these compounds gave semi-log plots with no curvature; the curvature of the lines in figure 8 is, however, definite. The result for cadmium silicate agrees well with that of Johnson and Davis (1939).

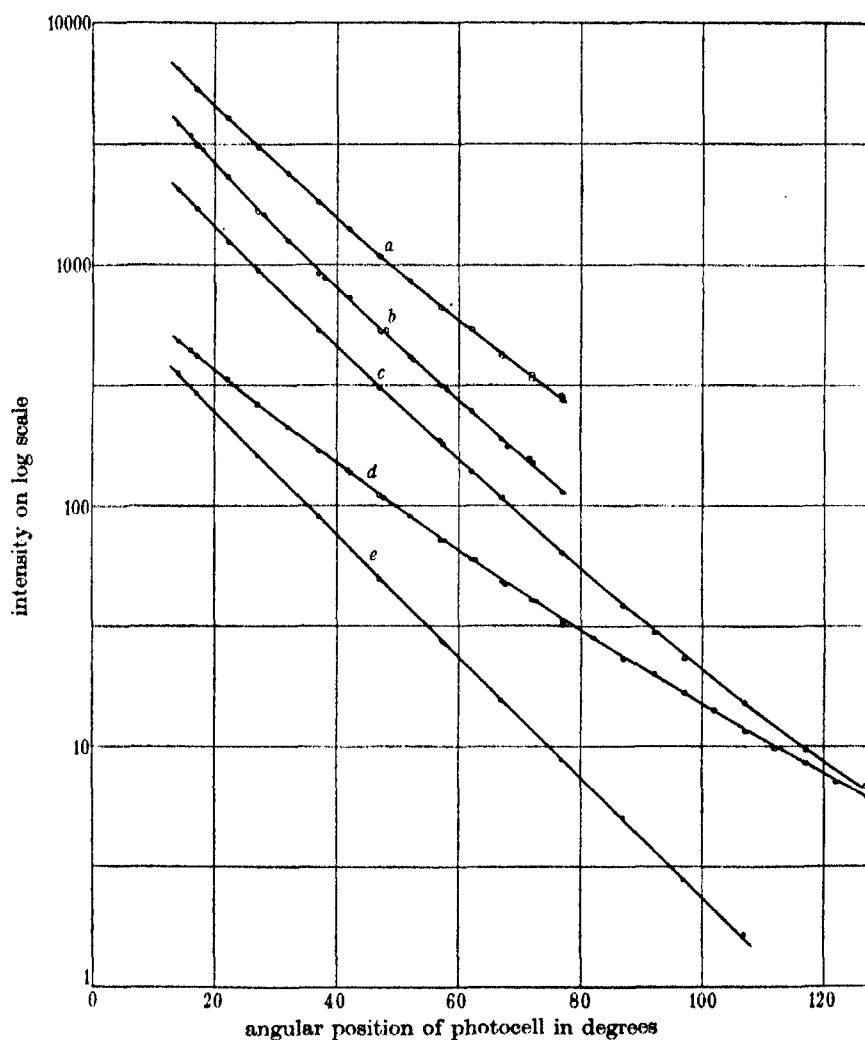


FIGURE 9. Decay curves. Curves *c*, *d*, *e* are composite curves formed by fitting together two partially overlapping curves. For curve *a*, $1 \text{ msec.} = 0.75^\circ$; for curves *b*, *c*, *d*, $1 \text{ msec.} = 1.5^\circ$; for curve *e*, $1 \text{ msec.} = 12^\circ$. Decay began at 3° .

- (a) Cadmium silicate, $1/\alpha = 25$ milliseconds at 40 milliseconds.
- (b) Zinc mesodisilicate, $1/\alpha = 10.9$ milliseconds at 40 milliseconds.
- (c) Cadmium chlorophosphate, $1/\alpha = 11.1$ milliseconds at 40 milliseconds.
- (d) Cadmium borate, $1/\alpha = 15$ milliseconds at 40 milliseconds.
- (e) Ammonium uranyl phosphate, $\alpha = 703 \text{ sec.}^{-1}$.

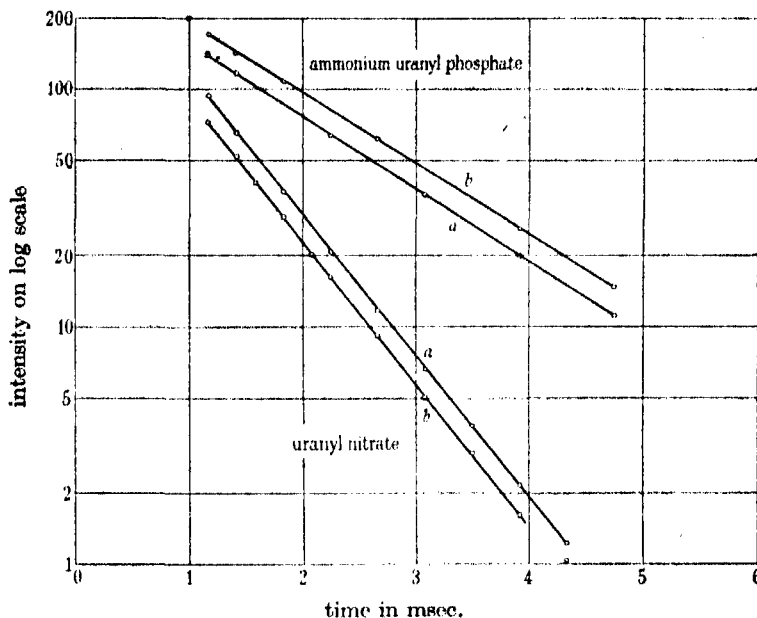


FIGURE 10. Uranyl salt decay curves at different exciting intensities. The exciting intensity for curves *a* is ten times that for curves *b*. Ammonium uranyl phosphate, $\alpha = 703 \text{ sec.}^{-1}$. Uranyl nitrate, $\alpha = 1390 \text{ sec.}^{-1}$.

TABLE 1

Phosphor	Mol. % of manganese
Cadmium silicate	0.5 in 10^3
Zinc mesodisilicate*	1 in 10^3
Cadmium chlorophosphate	1 in 14
Cadmium borate	3 in 10^3
Zinc orthosilicate	1 in 10^3 to 1 in 10^4
Zinc beryllium silicate	0.5-2 in 10^3

* Zinc orthosilicate is not, as stated by Fonda (1939), the only compound of zinc oxide and silica. X-ray diffraction measurements, together with luminescence results, clearly prove the existence of a mesodisilicate of composition ZnSi_2O_5 (Randall 1936).

As will appear from the discussion which follows, it is important to know whether the slope of such curves at any instant of time is a function of time after the exciting light is cut off or of phosphorescence intensity. Figure 11 shows that the intensity of the exciting light does not affect the decay curves for zinc mesodisilicate cadmium chlorophosphate and cadmium borate. The form in which the results are presented in figure 11 makes it clear that the slope of the curve is a function of the time only. It will be noted that the result for cadmium silicate is not very definite.

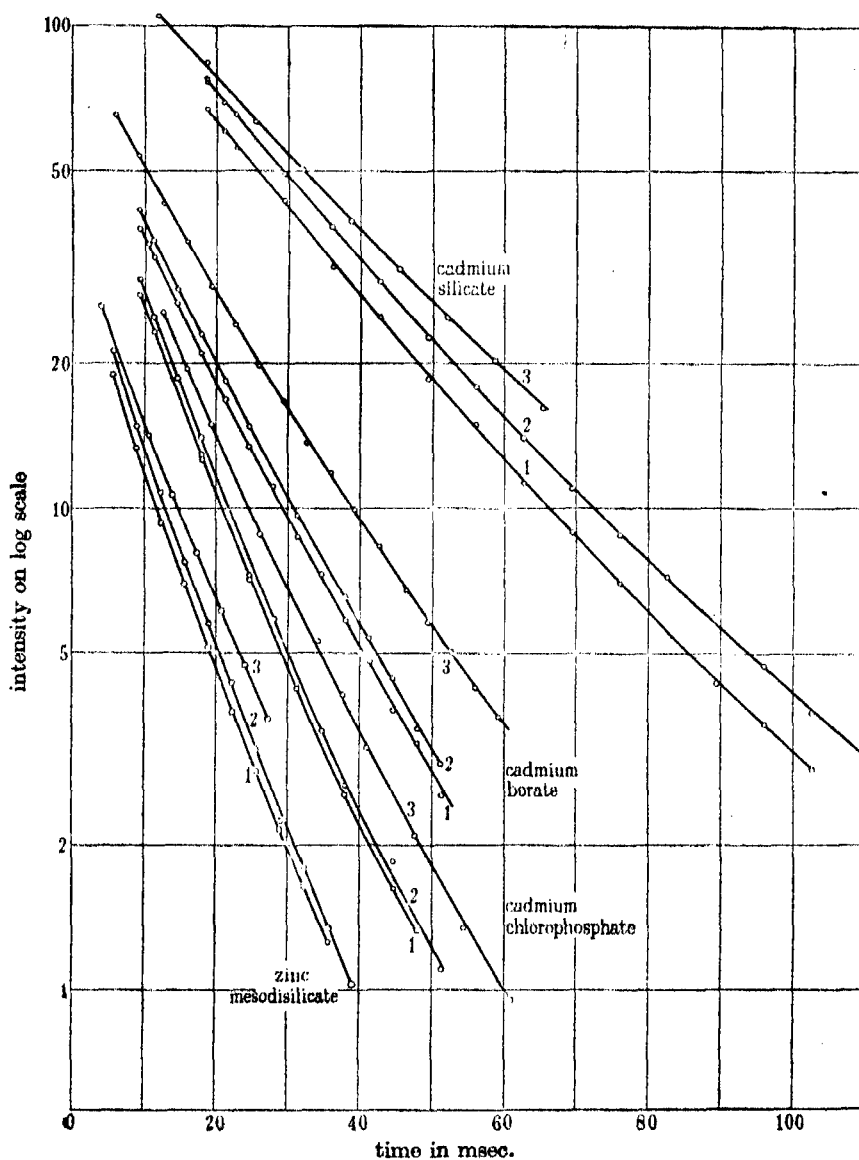


FIGURE 11. Effect of exciting intensity on decay curves of cadmium silicate, cadmium borate, cadmium chlorophosphate and zinc mesodisilicate. Curves marked 1 are for high exciting intensity, 2 are for low exciting intensity. In curves marked 3 the tails of the high intensity curves are shifted so that the actual phosphorescence intensity at a given time abscissa is equal to that of the low intensity curve. The fact that 1 and 2 are parallel but at an angle to 3 shows that the slope of the phosphorescence curve is a function of time and not of intensity.

In the case of some artificial willemites and zinc sulphides the initial phosphorescence decay is followed by a prolonged glow, and on the phosphoroscope disk the phosphorescent annulus glows through the whole 360°. This phenomenon was not exhibited to any extent by the compounds of table 1 other than the zinc orthosilicates.

The deviation from a straight line of the semi-log plots of the decay curves for the first four solids of table 1 was not great; but in the case of zinc beryllium silicate (see McKeag and Randall 1938 for further details of this type of phosphor) the deviation is greater, in fact the curve fits an exponential no better than it does a hyperbolic formula. The results are given in figures 12 and 13. The decay of zinc beryllium silicate, however, resembles the decays of other substances in that the curve is unaffected by intensity changes. This distinguishes this type of decay from that of zinc sulphide activated with copper or silver. The zinc beryllium silicate curves of figure 12 show also that the slope depends on the concentration of the manganese impurity.

In the case of zinc orthosilicates with manganese impurity (briefly referred to below as willemite) the various specimens which we studied all gave an initial decay curve which is approximately exponential and which is unaffected by exciting intensity (see figure 13). The curvature of the semi-log plot for the first 40 milliseconds of decay was always slight and about the same order as that for cadmium silicate or zinc mesodisilicate. In most specimens the slope of the curve decreases considerably after 40 milliseconds. Fonda (1939) states that this decrease of slope is temperature dependent and disappears at certain high and low temperatures. The decrease of slope did not occur (see figure 12) for one of our specimens (referred to as *A*) which had been subject to prolonged heat treatment and in which single crystals 0.2 mm. in size existed. Willemite with manganese contents from 0.1 to 1% composed of an agglomeration of crystals about 10 μ in size, have all much the same decay curve. All the specimens gave prolonged faint glows lasting several minutes, except *A*, which gave a fainter glow about the same intensity as that of a zinc beryllium silicate. The large crystals of specimen *A* were ground up and the fine powder separated with acetone from the coarse grains. The average size of the grains in the two samples was 10 and 200 μ , and the decay curves for the two samples are shown in figure 12. The decay curve is evidently independent of the grain size if the grains are split off large single crystals. Grinding the powder to a grain size of less than 1 μ or cooling rapidly from 1000° C fails to increase the long afterglow. A specimen of composition 1.8 ZnO, 0.2 BeO₂, 1.5 SiO₂ with 0.5% Mn gave a decay curve intermediate between those of ZnBeSiO₄ and Zn₂SiO₄.

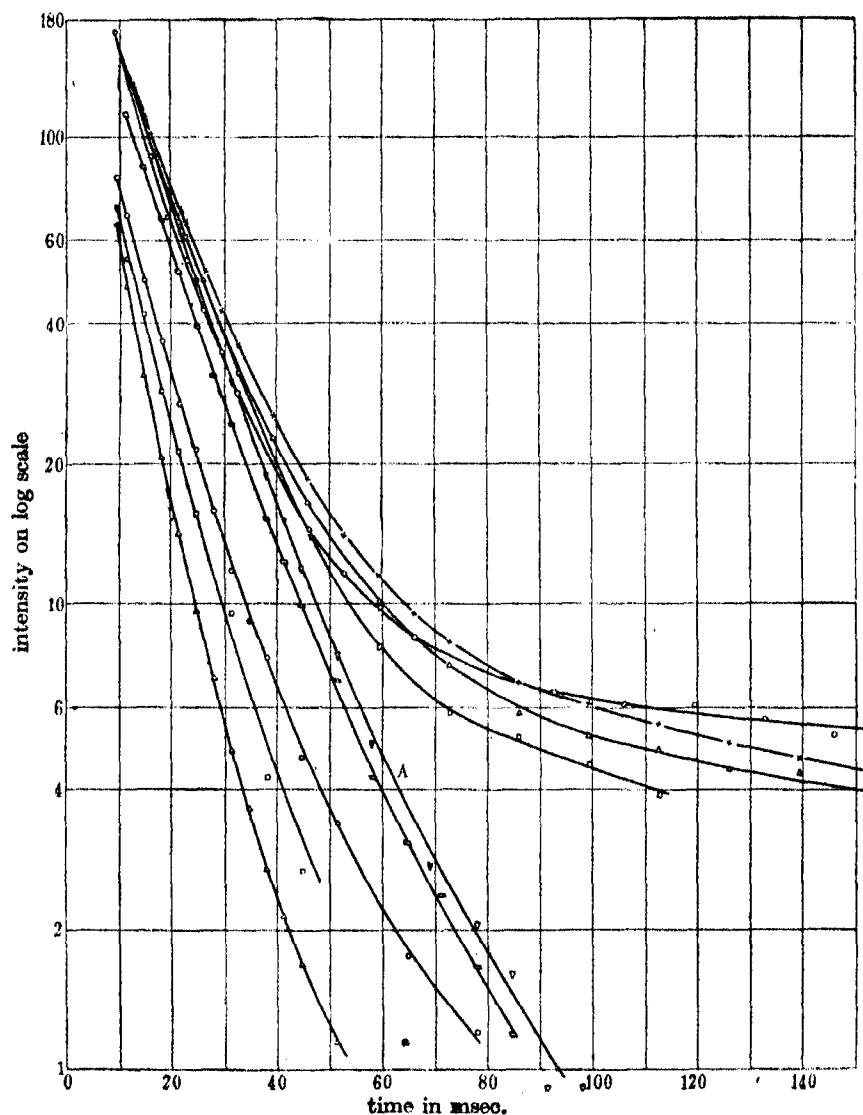


FIGURE 12

Willemities	$1/\alpha$ at 15 msec. in msec.
× 0.1% Manganese	14.3
□ 0.5% Manganese	13.1
△ 0.5% Manganese	13.4
○ 1.0% Manganese	11.4
▽ 0.5% Specimen A, 10 μ grains	12.3
◻ 0.5% Specimen A, 200 μ grains	12.8
Zinc beryllium silicates	
○ 1.8 ZnO, 0.2 BeO, 1 SiO ₂ + 0.5% Mn by wt.	11.0
□ ZnO·BeO·SiO ₂ + 0.1% Mn	10.1
△ ZnO·BeO·SiO ₂ + 2.0% Mn	7.6

The decay of zinc sulphide-Mn is of interest: the decay is independent of exciting intensity and temperature, and therefore is not of the bimolecular type (as is zinc sulphide-copper). We suggest that the decay process is essentially similar to that of the substances of table 1.

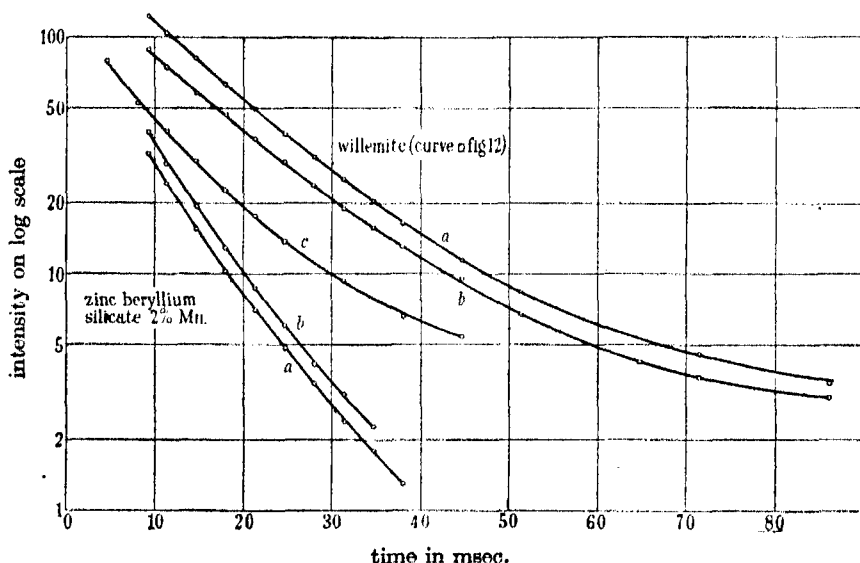


FIGURE 13. Willemite and zinc beryllium silicate decay independently of exciting intensity. Exciting intensity for curves *a* is ten times that for curves *b*. Curve *c* is the tail of the zinc beryllium silicate curve shifted as in figure 11.

PHOTOCONDUCTIVITY MEASUREMENTS

The fluorescent materials studied were available for the most part only as powders; as a result only qualitative measures of photoconduction could be obtained.

The powder was pressed gently in a layer about 0.2 mm. thick between quartz plates P_1 and P_2 (figure 14). A platinum sheet lying on the plate P_1 served as one electrode, while the other consisted of a series of fine graphite lines painted on the quartz plate P_2 . The light entered the powder by passing through the spaces between the graphite lines. Electrons produced by external photoelectric effect at the electrodes could not pass between the electrodes because the powder completely covered the electrodes. The arrangement was placed in a vacuum to reduce leakage currents due to moisture in the powder. Currents down to 10^{-14} amp. were measured using

an electrometer triode as amplifier. The ultra-violet light was produced by a monochromator and mercury arc.

Potentials of 200 V were applied between the electrodes. When the light was shone on the powder a current was produced instantaneously; this current then died away to zero roughly according to an exponential decay. This decay could be explained if the photoelectrons could not pass freely between the electrodes but accumulated as a space charge at for instance the surfaces of the powder grains. (Alternatively the space charge may be due to the fact that the positive charges cannot move and electrons cannot enter the powder from the cathode.) The rate of current decay calculated according to such ideas agreed with that found experimentally. The field in the crystal grains was due to the combined effect of the space charge and the applied field, but as the amount of space charge depended in an indefinite way on the previous applied fields and times of illumination of the powder, the currents measured are probably significant only to about 100%.

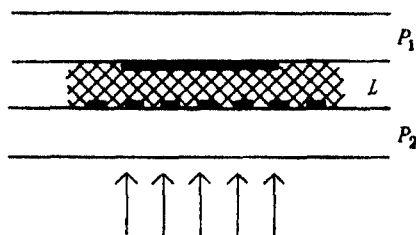


FIGURE 14

Tables 2 and 3 summarize the measurements of the initial photocurrents using two different wave-lengths of light. In both tables the following symbols are used to indicate the degree of fluorescence: ff, brightly fluorescent; f, faintly fluorescent; vf, very faintly fluorescent; nf, not fluorescent. The uranyl salts (table 3) fluoresce brightly but give no detectable photocurrent. All the manganese impurity phosphors give currents of about the same order. It should be noticed that the presence of manganese does not appear essential for the existence of photoconductivity; also it may be significant that photoconductivity is produced by long wave-length ultra-violet which excites negligible fluorescence (although Hofstadter (1938) obtained results for willemite which showed relatively very little photoconduction for these long wave-lengths). Calcium tungstate fluoresced brightly and gave a small but definite photocurrent.

TABLE 2. PHOTOCONDUCTION CURRENTS. PHOSPHORS CONTAINING MANGANESE AND THE CORRESPONDING PURE MATRIX. CURRENTS IN 10^{-12} AMP.

Phosphors with manganese impurity			Pure compounds		
Phosphor	Exciting wave-length		4360 A	Exciting wave-length	
	2650 A	3650 A		2650 A	3650 A
Zn_2SiO_4 (0.5 % Mn)	60 ff	40 vf	5 nf		
Zn_3SiO_4 (0.5 % Mn)	40 ff	15 vf			
Mineral willernite	20 ff	35 ff	Zn_2SiO_4	15 nf	20 nf
ZnBeSiO_4 (2 % Mn)	9 ff	3 nf	ZnBeSiO_4	0.4 nf	0.2 nf
CdSi_2O_5	12 ff	— nf	CdSi_2O_5	2.0 nf	0.05 nf
$3\text{Cd}_3(\text{PO}_4)_2 \cdot \text{CdCl}_2$	40 ff	25 nf	$3\text{Cd}_3(\text{PO}_4)_2 \cdot \text{CdCl}_2$	150 nf	150 nf
CdB_2O_4	75 ff	100 nf	CdB_2O_4	25 nf	100 nf
ZnSi_2O_5	10 ff	5 vf			
ZnS (Mn)	100 ff	1000 ff	100		
ZnS (Ag)	50 ff	2000 ff			
CaWO_4 (1 % Sm)	0.2 ff	0.2 ff			
Approx. ratios of energy in different wave-lengths	50	100	50		

TABLE 3. PHOTOCONDUCTION IN OTHER PURE SOLIDS

Currents in 10^{-12} amp.

Compound	Current using Hg 2560 A	Current using Hg 3650 A
Uranyl acetate	~ 0.02 ff	~ 0.01 ff
Uranyl nitrate	—	~ 0.01 ff
Calcium tungstate	0.2 ff	0.3 f

DISCUSSION

The sharp energy levels of single isolated atoms are in general broadened out into bands by the aggregation of the atoms into a solid. Such bands may or may not retain the characteristics of the atomic levels from which they derive. According to the theoretical work of Peierls, Slater, Mott, Seitz and others the forbidden energy levels lying between the possible zones of energies may contain isolated levels, known as excitation states. These ideas are represented diagrammatically in figure 15. Bands *A* and *C* contain possible energy levels for electrons, and band *A* in an insulating solid is full of electrons. Band *C* is generally referred to as the conduction band. The excitation states lying below the conduction band are denoted by *E*. We will now consider the bearing of these ideas on the experimental results just described.

If light shines on a crystal and is absorbed so that electrons are raised from the full band *A* into the excitation states *E* such electrons are still bound electrons and are not "free". In consequence no photoconduction can arise as the direct result of absorption into excitation states. Absorption into excitation states can however result in a radiation transition from the excitation state to the ground state. Such a transition from one of the states *E* to band *A* in figure 15 would be a luminescence transition. The extent to which the luminescence spectrum is of the line or band type depends on how much the levels *E* have been broadened. The probability that in time dt an electron returns to the ground state with the emission of light is given by $\alpha n dt$ where α is a constant for the crystal at a given temperature, and n is the number of electrons in excited states at time t after the switching-off of exciting radiation (see figure 1).

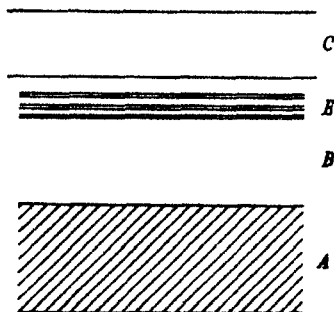


FIGURE 15. Energy levels in solids.

For this decay

$$dn/dt = -\alpha n$$

and

$$n = n_0 e^{-\alpha t}$$

or

$$I = I_0 e^{-\alpha t}, \quad (i)$$

where I is the phosphorescence intensity at time t .

The results for uranyl salts show the law of decay to be of this type with $\alpha \sim 10^{-3}$ sec.; the lack of photoconduction shows that in these salts we are dealing with a case of excitation states. It appears probable from the values of α that the transitions are of the "forbidden" type. Much faster decays ($\sim 10^{-8}$ sec.) would be expected for allowed transitions. The oscillograph records of figure 5 show that the value of α is for the uranyl salts almost temperature independent between room temperature and 90° K. Further confirmation that the decay for uranyl salts is simply exponential is obtained from the intensity curves of figure 10.

The phosphorescence of most of the solids dealt with above however depends on the presence of manganese impurity. Using the model of figure 15 it has been supposed generally that the presence of the impurity gives rise to new energy levels some of which may lie in band *B*. Absorption of light could then give either the exponential type of decay already observed for pure solids, or a hyperbolic type of decay if the electrons are raised to the conduction band *C*. The hyperbolic decay law arises as follows. Supposing that there are n electrons per unit volume in the conduction band, then there will also be n vacant impurity levels or positive holes in a filled band. The variation of n with time will now depend on the number of holes as well as on the number of electrons in band *C* and we shall have

$$dn/dt = -An^2.$$

Solution of this equation gives

$$I = I_0/(1 + An_0t)^2. \quad (\text{ii})$$

(This expression is characteristic of a bimolecular reaction.) From this we see that $1/\sqrt{I}$ is a linear function of time and the slope of the decay curve is proportional to \sqrt{I} . A further consequence is that such solids should be photoconductors.

Let us now examine these points for the solids cadmium silicate, zinc mesodisilicate, cadmium borate and cadmium chlorophosphate. Variation of exciting intensity over a wide range showed that the rate of decay was a function of time only and not of exciting intensity. This would immediately suggest a decay of exponential type similar to equation (1) above, the radiation transitions taking place from excitation levels of manganese ions to the ground state. The slight curvature of the semi-log plots could then be explained as a result of manganese ions having different environments in different parts of the crystal. This would lead to different values of transition probability and the law of decay would be of the form $I = \Sigma I_0 e^{-\alpha t}$. Such an arrangement would give a decay curve of the type observed and the decay would be independent of the exciting intensity. The rate of decay would however vary with the time of excitation. Our experimental arrangement did not allow wide variation of time of excitation and this point could not be checked. This explanation does not however take account of the photoconductivity of the specimens. While it is impossible to be certain, we are inclined to the view that the photoconduction arises in these solids from secondary causes. Reference to table 2 will show that the addition of quite large amounts of impurity makes comparatively little change in the photoconduction, and it may be that only a small proportion

of the excited electrons pass into the conduction band. A further point in favour of the view that the mechanism of luminescence in these solids is fundamentally one of excitation states in manganese ions is suggested by the fact that the mean slopes of the semi-log plots are not widely different in the four solids. Reference to figure 6 also shows the decay to be insensitive to temperature changes.

An alternative view is possible: the electrons may, on absorption of ultra-violet light by the solid, be set free from the manganese ions into the conduction band. If the recombination of these electrons by the manganese ions takes place in a time short compared with the life of the excitation state of the ion, the decay will still be essentially exponential and independent of exciting intensity.

The curvature of the semi-log plots for zinc beryllium silicate and zinc orthosilicate is much greater than that observed for the four solids considered above. The decay law is not, however, of a hyperbolic type, since the rate of decay is independent of exciting intensity. The initial part of the decay is approximately exponential and the mean slope is not very different from that of the manganese impurity phosphors already considered. The presence of a pronounced tail to the semi-log plot, especially in the case of zinc orthosilicate, makes it difficult to assume that excitation states only are involved. For electrons to remain trapped for a time of the order of minutes it seems necessary that they should first be freed into the conduction band and subsequently become trapped. It should be noted that in preparing such phosphors the solid is not usually kept at a high temperature ($\sim 1200^\circ\text{C}$) for more than an hour. In consequence it seems probable that at least some of the manganese impurity atoms will not have had time to reach positions of minimum energy. As a result the immediate surroundings will be unlike those of the normal lattice and a trap in the neighbourhood of an impurity atom is very probable. Thermal agitation of the lattice will be sufficient to eject electrons from traps of this or other types, and rapid recombination with impurity atoms will take place. In view of the fact that the slope of the initial position of the decay is for zinc orthosilicate roughly the same as that of other manganese containing phosphors we think it probable that the electrons are first captured into an excited state, and that the time constant of the initial decay is characteristic of the life of an excited state of the impurity atom. Figures 7 and 8 show that the decay is almost temperature independent, as would be expected if this view is the correct one. Fonda (1939), on the other hand, ascribes the initial exponential decay to the fact that at any moment in this period there are many more positive holes than free electrons.

This theory depends on the observation that the total number of electrons involved in the initial decay (~ 100 msec.) is much smaller ($\sim 1/50$ th) than the number involved in the temperature-dependent tail. Our alternative suggestion does not rest on this assumption. Moreover, it should be noted that the impurity produces little effect on the photoconductivity. Assuming values of 10^{-15} cm.² for the collision cross-section of the impurity ions, and a thermal velocity of 10^7 cm./sec. for the free electrons, the initial decay would be complete in a time of $\sim 10^{-10}$ sec. which is several orders of magnitude less than the value observed. Thus it appears that the current explanation (given above) of the hyperbolic decay, equation (ii), is not correct. While Fonda's theory would explain the exponential nature of the initial decay, and be in general agreement with our qualitative results on photoconductivity, it would not account for the observed time constant.

The results for tungstates are of interest in showing that these compounds are slightly photoconducting when fluorescing. There is no reason to suppose that the observed photoconduction current is a result of fortuitous impurity, and it seems probable that absorption takes place in the tungstate ion WO_4 and electrons are set free. The smallness of the current observed may be a result of rapid recombination.

We are greatly indebted to Professor M. L. E. Oliphant for the provision of facilities for this work, and to the Warren Research Fund Committee of the Royal Society for apparatus and a grant to one of us (M. H. F. W.). We also wish to thank Professor R. Peierls for much helpful discussion, and H. C. Cole for his valuable assistance in the construction of the apparatus and in the taking of observations.

SUMMARY

Observations of the laws of phosphorescence decay have been made on a number of solids, and an attempt has been made to correlate the results with associated measurements on photoconductivity. Pure uranyl salts decay exponentially and are not photoconductors. The luminescence of these salts is a property of excitation states in the co-ordination group of the uranyl ion, and the transitions are of the forbidden type.

The decay of the solids cadmium chlorophosphate, zinc mesodisilicate, cadmium silicate and cadmium borate is approximately exponential, and independent of phosphorescence intensity. The photoconduction of these solids is also present in the pure state and does not appear to be associated with the luminescence process. The phosphorescence of zinc orthosilicate

and of various zinc beryllium silicates is also approximately exponential initially, but subsequently slows down into a phosphorescent tail. It would appear that the initial decay is (as for the other solids with Mn impurity) determined by the lifetime of the excitation state of the manganese ion; the long tail process produces a number of free electrons, of which a certain proportion is subsequently trapped, probably in the neighbourhood of the Mn ions. The free electrons which escape trapping combine with the Mn ions very rapidly, forming excited Mn atoms; the time in the excited state is by comparison considerable, so these electrons are responsible for the initial exponential decay. The trapped electrons are slowly released by thermal vibrations and then combine with further Mn ions, giving rise to the phosphorescent tail.

REFERENCES

- Fonda 1939 *J. Appl. Phys.* **10**, 408.
de Groot 1939*a* *Trans. Faraday Soc.* **35**, 85.
— 1939*b* *Physica*, **6**, 275.
— 1939*c* *Physica*, **6**, 393.
Gudden and Pohl 1921 *Z. Phys.* **4**, 206.
Hofstadter 1938 *Phys. Rev.* **54**, 864.
Johnson 1939 *Phys. Rev.* **55**, 881.
Johnson and Davis 1939 *J. Opt. Soc. Amer.* **29**, 283.
McKeag and Randall 1938 British Patent, 480,356.
Nichols and Howes 1919 *Publ. Carneg. Instn*, no. 298, p. 38.
Randall 1936 British Patent, 457,126.
— 1939 *Trans. Faraday Soc.* **35**, 2.
Randall and Wilkins 1939 *Nature, Lond.*, **143**, 978.
Schleede and Bartels 1938 *Phys. Z.* **39**, 936.
Strange 1939 *Trans. Faraday Soc.* **35**, 95.
Warren 1935 *G.E.C. J.* **6**, no. 2 (May).
Wawilow and Lewschin 1928 *Z. Phys.* **48**, 397.
Winch and Machin 1935 *G.E.C. J.* **6**, no. 4 (Nov.).
-

Phosphorescence and electron traps

I. The study of trap distributions

BY J. T. RANDALL (*Warren Research Fellow*) and M. H. F. WILKINS
Physics Department, The University, Birmingham

(*Communicated by M. L. Oliphant, F.R.S.—Received 6 November 1940*)

Phosphorescence and electron traps

I. The study of trap distributions

BY J. T. RANDALL (*Warren Research Fellow*) and M. H. F. WILKINS
Physics Department, The University, Birmingham

(Communicated by M. L. Oliphant, F.R.S.—Received 6 November 1940)

The fundamental connexion between thermoluminescence, phosphorescence and electron traps in solids has been investigated. Thermoluminescence and long-period phosphorescence arise from the release of electrons from metastable levels or traps. By studying the thermal stability of trapped electrons and the probability of release from traps of different depths, methods have been developed for finding the depths of electron traps in phosphors. The main experiment consists in exciting the phosphor at low temperatures until all the traps are filled; the phosphor is then warmed at a steady rate and the light emitted while warming is measured as a function of the temperature. The results show that the trap distribution in impurity phosphors such as willemite and the alkaline earth sulphides are, in general, complex, and extend over a range often as wide as 0.2–1.0 eV. The probability of release of an electron from a trap of depth E at temperature T is of the form $se^{-E/kT}$, where s is a constant. Values of s for alkaline earth and zinc sulphides are in the neighbourhood of $10^8 \pm 1$ sec.⁻¹.

INTRODUCTION

As an introduction we shall give an outline of those ideas which are now generally accepted concerning the nature of a phosphor. The basis of a phosphor is a pure insulating crystal which is made luminescent by the addition of a small proportion of impurity atoms; these additional atoms are supposed to occupy interstitial or substitutional positions in the matrix lattice. The energy states of electrons in the crystal may be considered, partly, as being derived from states which exist in the individual atoms or ions which compose the crystal. The narrow energy levels which exist in the ions in free space are, in the lattice, broadened into bands (A and B , figure 1) by the field due to surrounding ions. These bands in an insulator are filled; and although the electrons in these bands may move fairly freely from ion to ion, the total state of motion cannot be altered by an electric field, and electrical conduction cannot take place. Some distance above the filled bands there is an empty band C corresponding to the states of an electron freed from an ion; this band is called the conduction band, because, if an electron is raised from B to C , the movement of the electron in C and

of the positive hole left in *B* can give rise to a current. Impurity atoms or lattice irregularities give rise to localized electron states (such as *D*, *E*, and *F*) with narrow energy levels which may occur between the energy bands of the pure crystal. An electron may be raised into the conduction band from *B*, *D*, or *F* by light absorption: this is the case of photoconduction; but if the electron is not given sufficient energy it may remain bound to its positive hole in an excited state at *E*, and then no photoconduction is observed. In the case of semi-conduction, the levels *F* are only a small way below the conduction band and electrons from *F* may be excited by thermal motion into the conduction band. In a phosphor the excited electrons return to the ground state at luminescence centres; the electron dropping

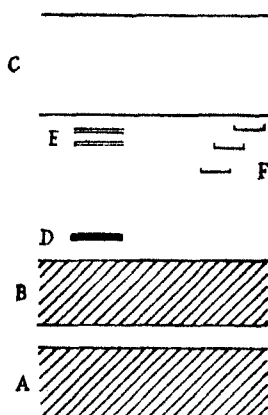


FIGURE 1. Energy levels in a phosphor.

first from *C* to *E* and then from *E* to *D*; the second transition being associated with the radiation of luminescent light. The levels *F* are normally empty, and are called traps, and may capture electrons excited into *C*. The release, by thermal agitation, of electrons from these trap levels gives rise to a delayed luminescent emission called phosphorescence. Also short period phosphorescence (not dealt with in this work) lasting not longer than a few milliseconds may arise owing to the finite time an electron spends in the excitation level *E*.

All non-metallic solids appear to contain some electron traps and many may be made to phosphoresce to some extent if suitably excited. In the case of alkali halides, the traps, called colour centres, have been studied in great detail by Pohl and his school (Pohl 1937); and a satisfactory picture has been evolved of their nature by Mott and Gurney (1940), who believe

that colour centres exist at points in the crystal lattice where a negative ion is missing. Traps in silver halides play a role in photographic processes; and have in this connexion been studied by several workers (e.g. Berg 1939). The presence of luminescence centres in a solid makes possible new methods of studying traps. This paper deals with the thermal stability of trapped electrons and its relation to thermoluminescence and with methods of finding the distribution of the trap levels at different depths below the conduction band. A second paper relates this work to long-period phosphorescence, and provides a quantitative explanation of phosphorescence in terms of the trap distribution. Previous work on phosphorescence (e.g. Lenard, Schmidt and Tomaschek 1928, Ives and Luckeish 1911, Mulder 1938) has been extensive, but has not been very susceptible to theoretical interpretation. An exception is the work on phosphorescent potassium chloride (Bünger and Flechsig 1931 *a, b*, Seitz 1938), which substance, however, differs in several respects from the more usual type of phosphor. The chemical and physical nature of traps in phosphors is not yet known: further work is required in this direction. It appears likely, however, that the traps are independent of the luminescence centres and exist in the pure matrix crystal like colour centres in alkali halides.

THE STUDY OF THERMOLUMINESCENCE GLOW

The glow experiment

It is a well-known fact that if a phosphor is excited while cold and is then warmed, light is emitted while warming. Such emission of light has all the properties of thermoluminescence; though, in the past, the term thermoluminescence has been used only for light emitted when a solid is heated above room temperature. In the present work thermoluminescence emission is referred to briefly as glow. The glow experiment consists in measuring the light emission when a thermoluminescent solid is warmed at a uniform rate; and the curve obtained of light emission against temperature is referred to as the glow curve. It has been generally assumed (e.g. Johnson 1939) that glow is due to release of electrons from traps; the probability of escape from a trap increasing with temperature and the shallower traps being emptied at lower temperatures. We have found that a study of the glow curve provides a means of showing how the traps are distributed in depth.

The experimental method is now described. A thin layer of the powdered phosphor is spread on the surface of a small copper box. A suitable phosphor

layer is made by putting a slight smear of glycerol on the copper and shaking the powder on. A copper-constantan thermocouple is soldered to the surface of the box, and an electric heater coil is arranged inside. Figure 2 shows two types of apparatus which have been used. Apparatus 1,

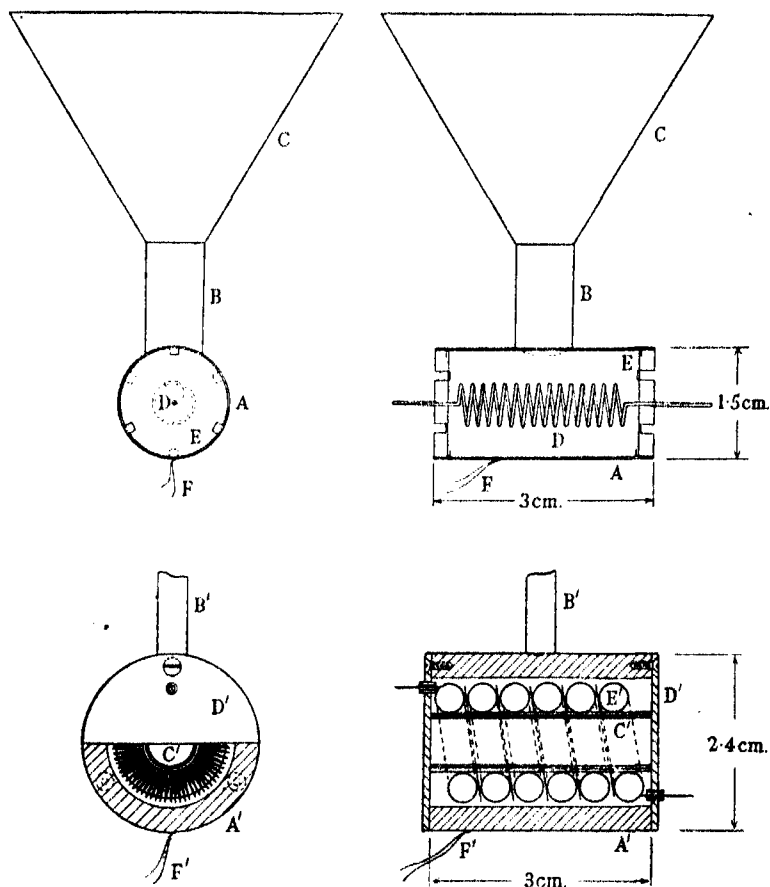


FIGURE 2. Apparatus

Apparatus 1

- A*, copper tube 0.2 mm. thick
- B*, german silver tube 0.05 mm. thick
- C*, copper funnel 0.04 mm. thick
- D*, heater coil 5 Ω , 6 V
- E*, mica discs 0.2 mm. thick
- F*, thermocouple leads 0.1 mm. diam.

Apparatus 2

- A'*, copper tube 3.2 mm. thick
- B'*, german silver tube 0.2 mm. thick
- C'*, copper tube 0.5 mm. thick
- D'*, copper end-plates
- E'*, coiled helical heater*
- F'*, thermocouple leads

* In *E'* the turns are separated by thin mica and insulated from *D'* by quartz tubes and from *C'* by asbestos. Coil resistance 100 Ω , 100 V.

in which liquid air is poured into the funnel *C*, is very convenient to use; but the rate of warming varies considerably with the temperature. Apparatus 2 was designed so that heat losses from the surface of the box would not affect the rate of warming, and so that high temperatures could be reached. For cooling, the whole system is placed in a beaker full of liquid air.

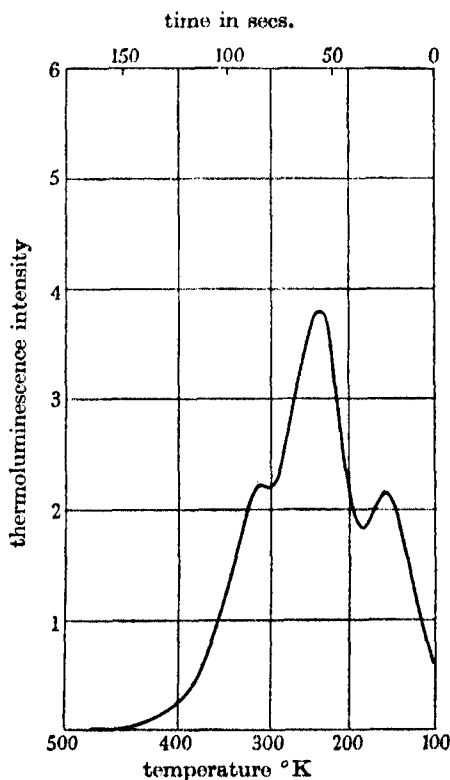


FIGURE 3. Glow curve for ZnS-Cu phosphor no. 2.

When the box is cooled with liquid air, the phosphor is excited with a mercury arc until all the traps are filled. The arc is then switched off, and the apparatus is placed in the dark in front of an electron multiplier type of photocell. The current in the heater is then switched on (in the case of apparatus 1 this is not done until the liquid air in the funnel *C* has all boiled away). The glow produced during the warming is observed as the deflexion of a galvanometer connected to the multiplier.

The galvanometer light spot is followed by a pointer connected to a pen which writes a trace on paper on a rotating drum. Marks are made on the

trace every 10 sec., and the thermocouple current is noted at the same time. The 10 sec. times are given by an electric bell connected to a clock. For accurate work two observers are required, one to follow the glow curve and the other to note the thermocouple readings. If the speed of the drum is constant the clock and bell can be dispensed with, and marks made on the

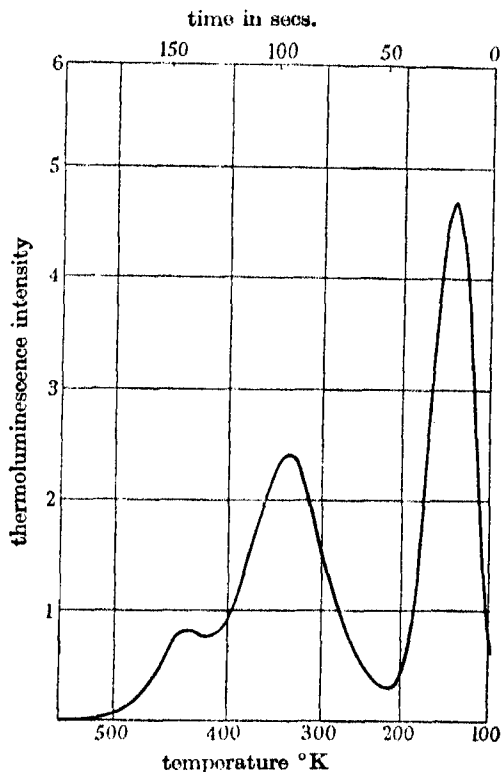


FIGURE 4. Glow curve for SrS-Bi phosphor no. 1.

trace when the thermocouple current takes certain values. (To avoid displacement of temperature readings relative to the glow curve, it is best to use the same period and damping for both the galvanometers employed.)

Glow curves of common phosphors are shown in figures 3 and 4. Before we begin to discuss the experimental results it is necessary to consider in more detail the theory of the form of the glow curve and the theory of some of the properties of trapped electrons.

THEORY OF THE GLOW CURVE

If the energy level of a trapped electron is E eV below the conduction band, the electron must absorb at least energy E before it can escape from the trap. By analogy with the case of a gravitational field acting on a ball in a hole, E is called the trap depth. The electrons in the traps have a Maxwellian distribution of thermal energies; hence the probability p of an electron escaping from a trap of depth E at temperature T is of the form

$$p = se^{-E/kT}, \quad (1)$$

where k is Boltzmann's constant and s is a constant which may, however, vary slowly with temperature.

If the trap is regarded as a potential box, s will express the product of the frequency with which the electron strikes the sides of the box and of the reflexion coefficient (cf. Mott and Gurney 1940, p. 136). There is, therefore, reason for supposing s to be of an order rather less than that of the vibrational frequency of the crystal, $\sim 10^{12}$ sec.⁻¹. In fact s is found to be $\sim 10^8$ sec.⁻¹ in the phosphors studied.

As we have said above, the glow experiment provides a method of finding how the traps in a phosphor are distributed with depth, since traps of a given depth provide a glow at a given temperature. In order that we may find the relation between the glow temperature and the trap depth, we will calculate the form of the glow curve when E is single valued and when the temperature rises at a constant rate.

Let n be the number of electrons in the traps at time t , then from equation (1),

$$dn/dt = -nse^{-E/kT}. \quad (2)$$

This assumes that there is no retrapping: that is, if an electron is liberated from a trap it always goes straight to a luminescence centre and does not on the way fall into another trap. We shall see in Paper II that this assumption of negligible retrapping is largely justified. The intensity of glow I is proportional to the rate of supply of electrons to the luminescence centres,

$$I = Cdn/dt = -Cnse^{-E/kT}. \quad (3)$$

Now from (2), $dn/n = -se^{-E/kT}dt$;

then writing $dT = \beta dt$, where β is the rate of warming, and integrating, we have

$$\log n/n_0 = -\int_0^T 1/\beta \cdot se^{-E/kT} dT,$$

and

$$n = n_0 e^{-\int_0^T 1/\beta \cdot s e^{-E/kT} dT};$$

hence

$$I = C dn/dt = n_0 C e^{-\int_0^T 1/\beta \cdot s e^{-E/kT} dT} s e^{-E/kT}. \quad (4)$$

This expression represents the glow curve for a phosphor containing traps of one depth. The curve has been calculated for $s = 2.9 \times 10^9 \text{ sec.}^{-1}$ and $E = 0.67 \text{ eV}$, and for two different rates of warming (see figure 5).

Beginning at low temperatures, the curve rises exponentially according to the exponential term in equation (3). When the light emission has

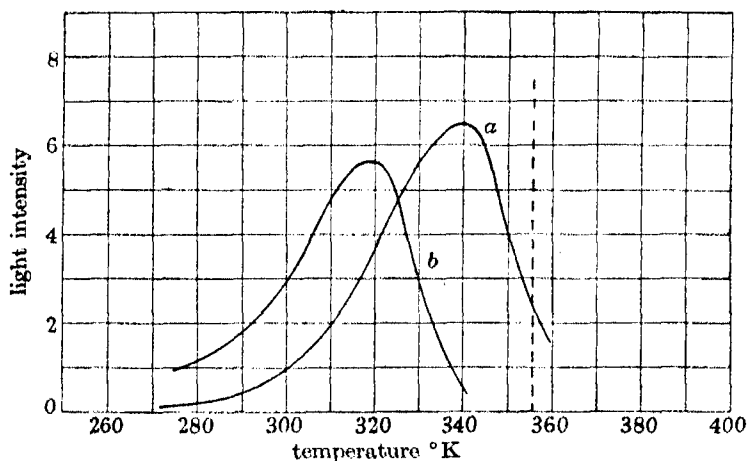


FIGURE 5. Theoretical glow curve for a single trap depth and for two rates of warming, (a) 2.5° per sec., (b) 0.5° per sec. $E = 0.67 \text{ eV}$; $s = 2.9 \times 10^9 \text{ sec.}^{-1}$; at 356° K the decay period of the trapped electrons is 1 sec.

continued for some time the number of trapped electrons (n in equation (3)) becomes appreciably diminished and the curve, after reaching a maximum, falls, and when all the traps have emptied reaches zero. The maximum intensity of glow occurs at a temperature somewhat below the temperature at which the probability of an electron escaping from the trap is 1 per sec. In figure 5 the probability is 1 per sec. at 356° K ; hence we may write

$$s e^{-E/kT_G(1+f(s, \beta))} = 1,$$

where T_G is the temperature of maximum glow and $f(s, \beta)$ has a value small compared with one.

Hence

$$E = T_G \{1 + f(s, \beta)\} k \log s. \quad (5)$$

The temperature of maximum glow varies only slightly with the rate of warming. We will refer to this temperature as the temperature which corresponds to the trap depth. From equation (5) we see that the trap depth is proportional to the corresponding temperature; so it is possible to establish a relation between trap depth in electron volts and temperature of glow in $^{\circ}\text{K}$, and thus to calibrate the thermal glow spectrum in units of trap depth. The above calculation shows that 0.67 eV corresponds to 340 $^{\circ}\text{K}$ at the rates of warming normally used. Hence the relation is approximately that 50 $^{\circ}\text{K}$ corresponds to 0.1 eV or that the trap depth is approximately $25kT$. Thus the glow near liquid air temperature (90 $^{\circ}\text{K}$) corresponds to trap depths of 0.2 eV. Figure 5 shows that the form of the glow curve is altered little by change in the rate of warming, but that the curve is slightly displaced as a whole towards lower temperatures when the rate of warming is decreased: this point has been experimentally confirmed.

In practice the trap depth E is not single valued but varies over wide ranges, and the glow curve consists of broad peaks built up from the overlapping peaks due to the various E values (see figures 3 and 4).

THE CONNEXION BETWEEN PHOSPHORESCENCE AND THERMOLUMINESCENCE

Phosphorescence and glow are terms used to describe different forms of the same phenomenon. In the case of phosphorescence the temperature is kept constant, but when glow is produced the temperature is rising rapidly. There is, however, a correspondence between the phosphorescence curve and the glow curve; for it is the deep traps which contribute both to the phosphorescence after long times and to the thermal glow at high temperatures, and similarly the shallow traps contribute to the short times of phosphorescence decay and to the glow at low temperatures.

A given temperature of glow corresponds to a given trap depth, and if phosphorescence takes place at a given temperature it is possible to calculate the mean time an electron stays in the trap of given depth.

From equation (4) we have seen that the trap depth E corresponds to a temperature of glow given by equation (5). If phosphorescence takes place at a constant temperature T , the mean time t that an electron spends in a trap is the reciprocal of its probability of release, and is by equation (1):

$$t = 1/p = s^{-1} e^{E/kT}.$$

Hence

$$\log t = E/kT - \log s,$$

and substituting for E from equation (5),

$$\log t = \log s \frac{T_G \{1 + f(s, \beta)\} - T}{T}, \quad (6)$$

where T_G is the glow temperature.

This expression provides the required relation between the time of phosphorescence decay t and the temperature of glow. It also shows at what rate different parts of the glow curve will decay if the phosphor is kept at a constant temperature.

EXPERIMENTAL RESULTS

1. *Traps in some phosphors activated by manganese*

All the powders in this group were excited by short wave-length ultra-violet from a quartz mercury arc. The phosphorescent properties have been described by Randall and Wilkins (1945).

The willemite powders (zinc silicate-Mn) give the brightest glow; figure 6*e* is typical. The various peaks can be observed qualitatively in a very convenient way. The powder is spread on a piece of copper foil which is dipped in liquid air and illuminated by ultra-violet radiation; the foil is then withdrawn and allowed to warm in the dark; the liquid air hanging on the bottom edge of the foil causes a temperature gradient to be formed in the copper; the glow peaks are observed as bands of light moving rapidly downwards across the foil. Different powders can be compared by spreading them side by side in vertical strips on the foil.

All willemites gave peaks at the same temperatures, except a fluorescent mineral specimen which gave no visible glow at all. There are two characteristic small peaks at 125° and 170° K; there is also a trace of a peak nearer liquid air temperature, but this peak decays rapidly and gives rise to phosphorescence at liquid air temperature. There is a larger peak with a maximum which varies in position from 245° to 260°; the height of this peak also varies considerably. By applying equation (6), we see that at room temperature the traps in the temperature range of the large peak give rise to phosphorescence of period ~ 0.1 sec. This is the temperature dependent phosphorescence tail observed by Fonda (1939). At temperatures much higher or lower than room temperature this phosphorescence disappears. The glow peak in question was almost entirely absent in specimen A (Randall and Wilkins 1945); and this accounts for the absence of a noticeable phosphorescence tail in this specimen. If the peak extends much

above room temperature a phosphorescence lasting seconds and minutes may be observed.

The zinc beryllium silicates (figure 6 *h, i*) appear to have a trap distribution similar to that of the willemites. The phosphorescence is similar.

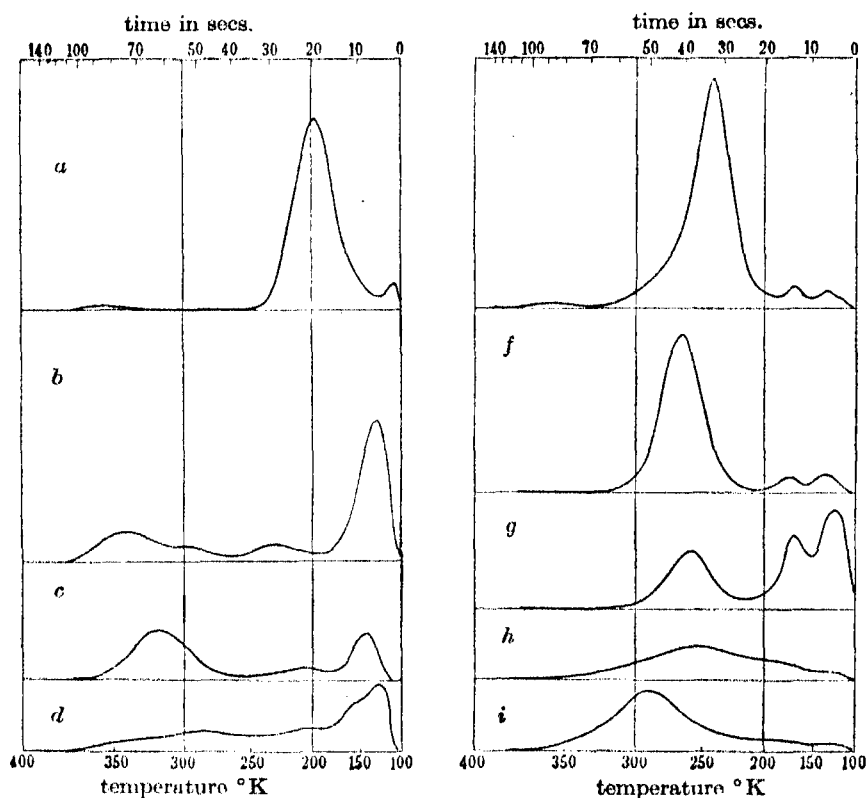


FIGURE 6. Thermal glow curves for phosphors activated by manganese. *a*, Cadmium borate. *b*, Cadmium chlorophosphate. *c*, Zinc mesodisilicate. *d*, Cadmium silicate. *e*, Green zinc silicate 0.5% Mn. *f*, Zinc silicate 1% Mn. *g*, Zinc silicate 0.1% Mn. *h*, Zinc beryllium silicate Mn. *i*, Zinc beryllium silicate 2% Mn.

Cadmium borate (figure 6*a*) has practically no room temperature phosphorescence, as there are no traps of the depth required to produce such phosphorescence. The peak near liquid air temperature gives rise to strong phosphorescence at liquid air temperature. Cadmium chlorophosphate, zinc mesodisilicate, and cadmium silicate (figure 6), all have slight room temperature phosphorescence. None of the powders in this group have more than a trace of a phosphorescence tail of period ~ 0.1 sec.

No general similarity is apparent in the glow spectra of the various phosphors activated by manganese; and these experiments give no indication that the traps are associated with the manganese impurity.

When studying the photoconductivity of these phosphors (Randall and Wilkins 1945), it was found that space charges were produced in the phosphor after the photoconductivity current had flowed for a time. This space charge decayed away slowly in the dark when there was no applied field. The most probable explanation is that the space charge is produced by electrons which travel through the crystal some distance and are then trapped. In the dark the electrons are slowly released from the traps, this process corresponding to phosphorescence, and the space charge is dispersed. The magnitude of the space charge shows that, at the lowest limit, there are 10^{12} electrons trapped per c.c. of phosphor. The space charge can be stored in both pure zinc silicate and zinc silicate activated by manganese; this indicates that the traps in this phosphor are a property of the crystal matrix and not of the manganese impurity. Infra-red light has little effect in dispersing the space charge in the phosphors of the group we are considering; and this agrees with the observation referred to in the section on the optical properties of trapped electrons—that infra-red has little effect in releasing the glow from these phosphors.

An important observation has been made by Herman and Hofstadter (1940). A zinc silicate-Mn was exposed at liquid air temperature to ultra-violet light and then warmed in the dark. A sudden increase in the dark current was observed at 213° K. This temperature corresponds to that of the large glow peak observed in our experiments, and it is almost certain that the electrons which contribute to the dark current are the same electrons which produce the thermal glow. Thus the observation shows that the electrons do pass through the conduction band during the phosphorescence process. Another indication of this is in the observation that a glow is produced in phosphorescent zinc silicate when an electric field is applied to the phosphor in the dark. The electric field apparently assists the electrons to leave the traps (cf. Frenkel 1938).

The observation made on the dark current of zinc silicate suggests a new kind of experiment in which the variation of dark current with temperature is measured during the glow experiment. In this way trap distributions could be studied in non-fluorescent solids. The experiment could also be used to distinguish between semi-conductors which contain equal numbers of electrons and semi-conduction levels and those which contain less electrons than levels. In the second type of semi-conductor the trap distribution could be found by the method, and this would supplement the

knowledge already gained from experiments in which the semi-conduction current is measured under equilibrium conditions at different temperatures. The experiments which are described in this paper, on trap distributions in phosphors, indicate that most impurity levels in solids are distributed in depth. The trap levels in phosphors resemble in many ways semi-conduction levels; it is, therefore, reasonable to suppose that the apparent variation with temperature of the activation energy in semi-conductors is to be attributed to a distribution with depth of the semi-conduction levels.

2. *The phosphorescence and thermal glow of potassium chloride activated with thallium*

The phosphorescence of KCl-Tl has been studied with great thoroughness by Büniger and Flechsig (1931*a* and *b*), and an interpretation has been given by Seitz (1938) and Hilsch (1937). The study of the glow, however, enables some further information about the phosphorescence to be obtained. Also KCl is of special interest in that specimens can be obtained with a single trap depth E , while in all other phosphors we have examined there is a wide distribution of E values in any one specimen. The single value of E should give rise to a glow peak of the type calculated above (figure 5); and hence, by using Büniger and Flechsig's values for E and s , it is possible to check the theory we have given of the glow curve, or alternatively to assume the theory and check Büniger and Flechsig's values.

Büniger and Flechsig found that in some specimens of KCl-Tl the phosphorescence (which is almost entirely in the ultra-violet) decays near room temperature almost exactly according to an exponential law. The rate of decay changed rapidly with temperature and the decay law fitted the formula:

$$I = I_0 \exp(-se^{-E/kT}),$$

where I is the intensity of phosphorescence, E is an activation energy equal to 0.67 eV and s is a constant $2.9 \times 10^9 \text{ sec.}^{-1}$

The specimen of KCl-Tl (kindly given us by Prof. Pohl) which we have used was not a specimen with a single E value; but it was found that the value of E was distributed only over a narrow range, the E values differing by less than 1%. This small distribution of E values, however, causes the phosphorescence decay law to differ considerably from an exponential form. It must be pointed out that it is not impossible that the constant s will vary as well as E , but experimentally it is difficult to distinguish between trap distributions in which E varies slightly and s is constant and those in which E is constant and s varies slightly. This distinction is not of much physical

importance, and for convenience of discussion we will assume s is constant and E varying.

Figure 7 shows the decay curves plotted on a logarithmic scale for KCl-Tl at various temperatures: a slice of crystal was excited by a mercury arc and the phosphorescence curve recorded by the multiplier and rotating drum.

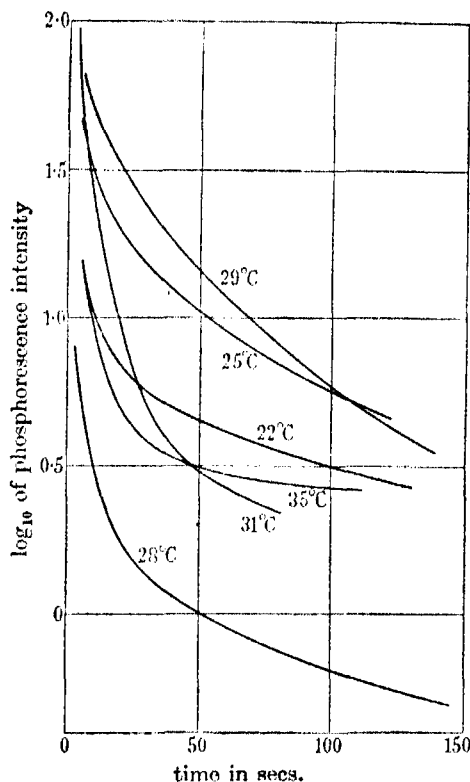


FIGURE 7. Phosphorescence decay curves of KCl-Tl at different temperatures.

The exciting light used for the 28° C curve was $\frac{1}{30}$ normal intensity.

We will now consider what distribution of E values will give rise to the phosphorescence observed. The E value will not be expressed in electron volts, but as the corresponding glow temperature.

For the rate of warming used (1.3° per sec.) equation (6) gives the following connexion between period t of phosphorescent decay at temperature T and the glow temperature:

$$\log_{10} t = 9.5 (T_0 \times 1.08 - T)/T,$$

taking s as $10^{9.5} \text{ sec.}^{-1}$

The results of figure 7 may be stated in the following way, the temperature of glow corresponding to the period of decay being calculated from the formula above.

Temperature of phosphorescence	Characteristics of the decay	Glow temperature corresponding to period of phosphorescence decay
295° K	Predominant period of decay is ~100 sec. With a lesser proportion of period ~15 sec.	336° K 310° K
302° K	Predominant period is ~30 sec.	328° K
308° K	Predominant period is ~10 sec. With a lesser proportion of period ~200 sec.	316° K 358° K

The glow peak should be built up from component glow peaks which have maxima at different temperatures corresponding to different E values. The mean temperature of the predominant period is, from the table, 327° K. Hence the predominant component of the glow peak should have its maximum at 327°; in other words, the maximum of the E distribution should be at 327°. The E distribution should on this scale of corresponding temperatures extend from 310° to 358°.

To test the correctness of these ideas the glow curve has been studied. Powdered KCl was used on the glow apparatus (type 1). The powder was excited at room temperature, a drop of liquid air was then placed in the funnel and warming was begun from below room temperature. The powder phosphoresces at an appreciable rate while at room temperature, and if this phosphorescence is allowed to proceed for some time before the glow curve is recorded, the resulting glow peak will be much smaller than the undecayed peak. Figure 8 shows the glow curve after different times of phosphorescence at room temperature. If the peak corresponded to a single value of E , the peak would decay as a whole and its maximum would remain at a fixed temperature. In fact the maximum moves to higher temperatures as the peak decays; which means that the components of the peak with maxima at lower temperatures decay more rapidly than the higher temperature components. From the data of figure 8 the E distribution can be deduced. For, roughly, the height of the E distribution curve at a certain temperature is proportional to the height of the decayed peak which has its maximum at that temperature. The dotted curve in figure 8 shows the approximate E distribution curve deduced in this way. The curve agrees with the rough estimates obtained from the phosphorescence data.

In the above we have shown that the phosphorescence decay curve may be correlated with the glow curve by assuming the value of s given by Bün̄ger and Flechs̄ig. This amounts in fact to a confirmation of that value of s . It is possible to make a more precise check of the E and s values given by Bün̄ger and Flechs̄ig (and alternatively to test the correctness of the theory given above of thermal glow) by comparing the experimental glow curve of figure 8 with that of figure 5 which has been calculated by assuming the E and s values of Bün̄ger and Flechs̄ig. When the different rates of warming for figures 8 and 5 are taken into account, the position of the maxima of the peaks is found to agree to 2%; and the width and shape of the peak also agrees well.

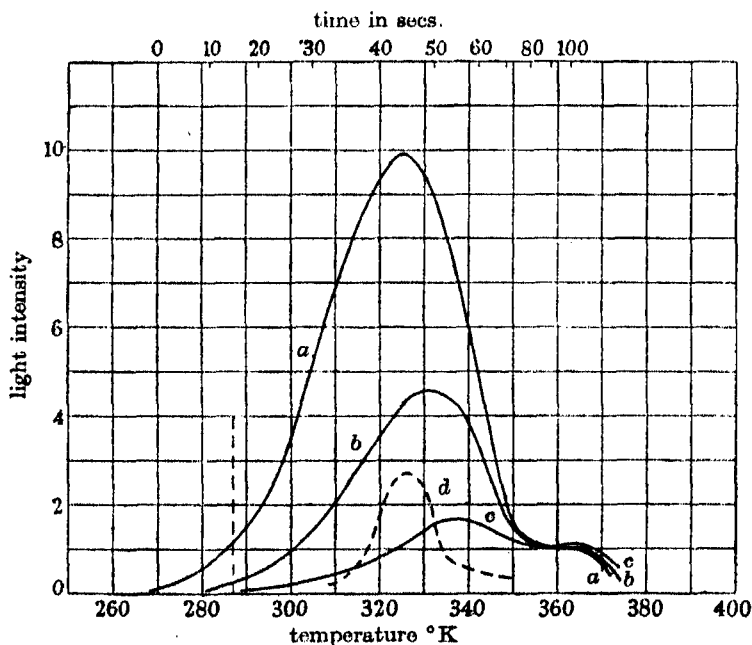


FIGURE 8. Glow peak of KCl-Tl after decay at 287° K. Time of decay: (a) 10 sec., (b) 7 min., (c) 33 min. Curve d shows the probable distribution of traps at different depths.

The discussion so far has referred only to one part of the E distribution, in the region of 0.7 eV or 330° K. The curves in figure 8 show that there are also traps in the region of 370° K. It is not possible to tell from the glow curve alone whether there is a sharply defined E level at 370°, or whether the E values are spread over a region several degrees wide; to use an optical analogy, the resolving power of the glow experiment is not

sufficiently great. To analyse the peak at 370° it would be necessary to measure the phosphorescence decay curve at a suitable temperature such as 340° , or to measure the peak after various times of decay at that temperature.

The whole glow spectrum from 90° K upwards is shown in the lower curves of figure 9. In all there are four peaks in the spectrum. The peak near 200° K would correspond at room temperature to a decay period of

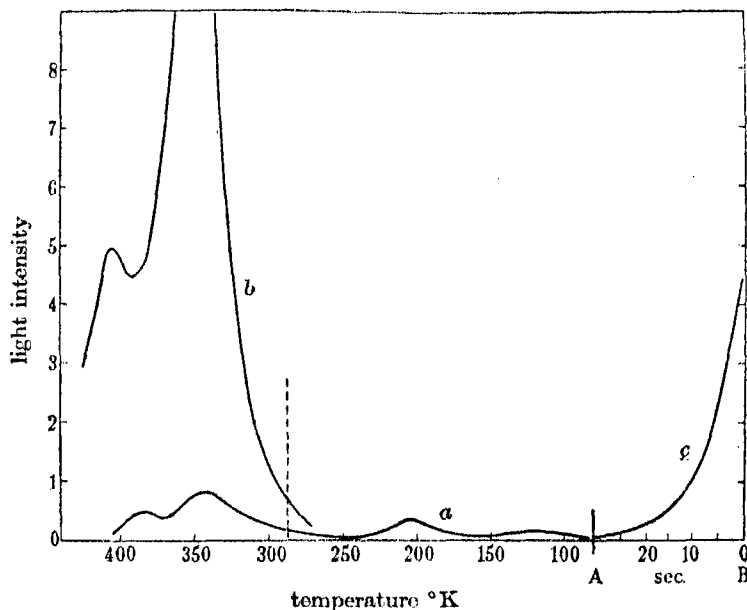


FIGURE 9. Glow of KCl-Tl when excited (a) at 90° K, (b) at 286° K. (c) is the phosphorescence curve at 90° K which is continuous with the glow curve (a). Thus the temperature is kept constant at 90° K from A to B and starts to rise at the point A. The mean rate of warming is about 3° per sec., hence the position of the high temperature peaks is slightly different from those in figure 8.

$\sim 10^{-3}$ sec. Bunger and Flechsig have, in fact, observed a very short decay period which they suggested was 0.5×10^{-4} sec. It may well be that this quick decay is due to release of electrons from the 200° K traps. There is also a glow peak near 100° K which causes intense phosphorescence at 90° K. The decay curve of this phosphorescence is almost exponential (see figure 9c), hence the E level is sharply defined.

The two curves, (a) and (b), in figure 9 illustrate a phenomenon which has not been observed in any phosphor other than KCl-Tl. When the phosphor

is strongly excited at 90° K, very little energy is subsequently released in the glow peaks between 300° and 400° . If excited at room temperature, about ten times as much energy is released from these peaks. To explain this phenomenon it must be remembered that KCl-Tl is an exceptional phosphor in that it does not photoconduct (Hilsch 1937), while all the other phosphors we have used are known to photoconduct (e.g. Randall and Wilkins 1945). Also the trap levels in KCl are much more sharply defined than those in other phosphors. In any ordinary phosphor the traps are almost certainly separate from the luminescence centres, and the electron wanders about the crystal from the fluorescence centre to the trap. In KCl-Tl, Seitz (1938) suggests that the trap consists of a metastable level in diatomic thallium molecules which are also the luminescence centres. The excited electrons do not at any time pass outside their particular centre. The different trap depths we have observed probably correspond to different levels in the thallium molecules. The coexistence of these various levels in one centre is also indicated by the structure in the infra-red quenching spectrum (Bünger and Flechsig 1931 *b*). In an ordinary phosphor which photoconducts there are, apparently, always more luminescence centres than traps; and hence at whatever temperature the phosphor is excited, all the traps corresponding to higher temperatures are filled and the phosphor is said to be saturated; this is shown by the fact that the saturated glow curve is the same height at whatever temperature excitation takes place. In KCl-Tl, however, for low temperature excitation each excited electron enters the shallower trap level in its particular thallium molecule and the deeper trap levels are not all filled. When excited at higher temperatures the electrons are not stable in the shallower traps, and the electrons all fill the deeper traps. This idea is supported by the observation that the total light energy which can be stored (in phosphorescence and glow) at 90° K is about equal to that which can be stored when excitation takes place at room temperature. This total of energy corresponds to the total number of thallium molecules in the phosphor.

*The determination of the value of the constant s by
measuring the decay of the glow curve*

As has been said before, the emission of energy during phosphorescence corresponds to a diminution in the amount of energy stored in the phosphor and to a diminution in the height of the glow curve. If the phosphor is kept at a fixed temperature T during phosphorescence, the various parts

of the glow curve, corresponding to different trap depths, decay at rates given by equation (6),

$$\log t = \log s \frac{T_G \{1 + f(s, \beta)\} - T}{T'},$$

where T_G is the glow temperature and t is the period of decay of that part of the glow curve.

These decay periods can be found by measuring the glow curve in various stages of diminution after phosphorescence has taken place for various times. Such measurements provide a method for finding a value for the constant s in phosphors which contain a wide distribution of trap depths. It must be noted that equation (6) is only approximate, as in its derivation it is implicitly assumed that each part of the glow curve corresponds to a single E value and to a single period of exponential decay. In fact, the glow curve is built up from component glow peaks of finite width, each peak corresponding to a single value of E . Hence, during phosphorescence, each temperature of the glow peak corresponds to a range of decay periods the mean of which is given by equation (6). It has been found that the most convenient way of finding s —and the way which seems least affected by the approximation contained in equation (6)—is to allow the glow curve to decay for various times, and to find the glow temperature at which the glow curve has fallen to half its undecayed value, and to plot these temperatures against the logarithm of the time of decay.

From equation (6) it is seen that the graph T_G against $\log t$ will be very nearly a straight line which makes an intercept on the $\log t$ axis, when T_G is zero, which is equal to $\log s$. Also, the line should cut the temperature axis, at a point T_1 , below the temperature at which decay took place. At this temperature T_1 , $\log t$ is zero; hence from (6)

$$f(s, \beta) = \frac{T - T_1}{T_1}.$$

Figure 10 shows the experimental results for a ZnS-Cu phosphor (no. 2). The decay took place at room temperature and the apparatus was cooled slightly before the glow curve was recorded. At high temperatures (near 440° K) the glow curve is not appreciably decayed, and the various glow curves overlap accurately. The undecayed glow curve was obtained by exciting the powder at a temperature well below room temperature. The amount of stored energy (the area under the glow curve) decreases rapidly with time; hence the phosphorescence decay of this type of phosphor is rather rapid.

Figure 11 shows representative decayed glow curves for a SrS-Bi phosphor (no. 1).

The decay of the glow curve of the SrS-Bi phosphor (no. 1) while phosphorescing at liquid air temperature is shown in figure 12. Frosting up of the powder while the decay took place was avoided by placing the powder (on apparatus 1) in a box with a cellophane window warmed by an air blast at about 100° C. Figure 12 shows that for long times of decay the glow curve decays in all temperature regions; this effect being superimposed

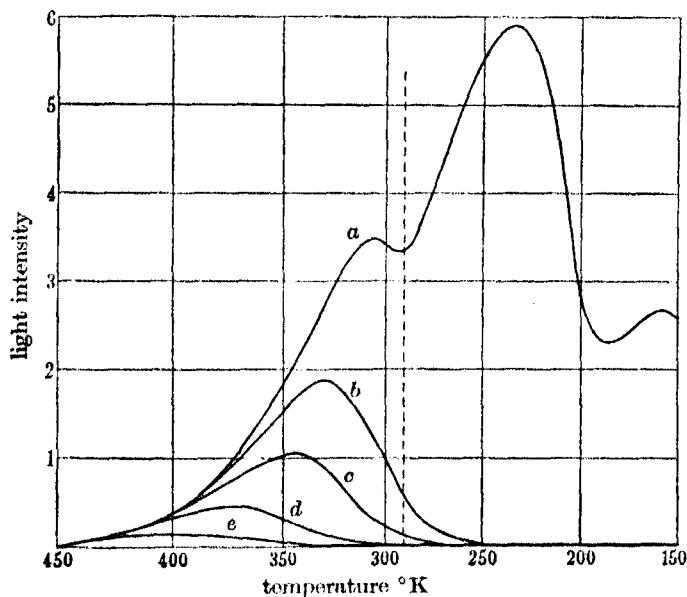


FIGURE 10. Glow curve for ZnS-Cu phosphor after various times of phosphorescence decay at room temperature 290° K marked by the dotted line. Rate of warming 2.2° per sec. Time of decay (a) zero, (b) 25 sec., (c) 140 sec., (d) 20 min., (e) 95 min.

on the decay of the curve on the low temperature side, as predicted by equation (6). This phenomenon is not general in phosphors and is probably caused by absorption of infra-red radiation of wave-length 2-3 μ , which may be radiated from the box and the warm window. In order that the decay predicted by equation (6) may be separated from the general decay of the glow curve, the decayed glow curve has been multiplied in height by a factor such that the height of the curve at high temperatures is the same as the height of the undecayed curve. The temperature at which the glow curve has decayed to half its initial height is measured from the curve multiplied in this way.

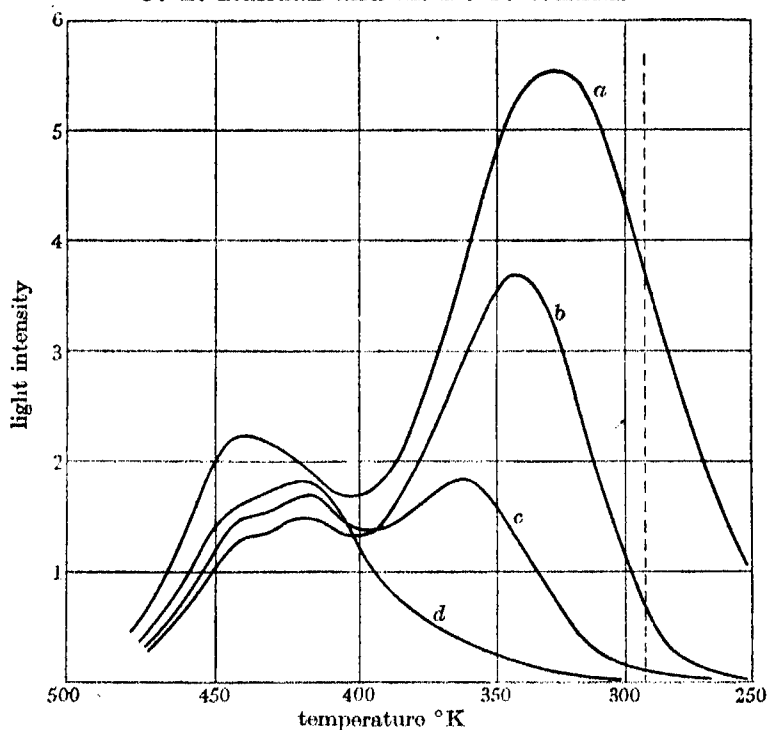


FIGURE 11. Glow curve for SrS-Bi (no. 1) powder after various times of phosphorescence decay at room temperature 293°K marked by the dotted line. Rate of warming 2.2° per sec. Time of decay (a) zero, (b) 35 sec., (c) 9 min., (d) 80 min.

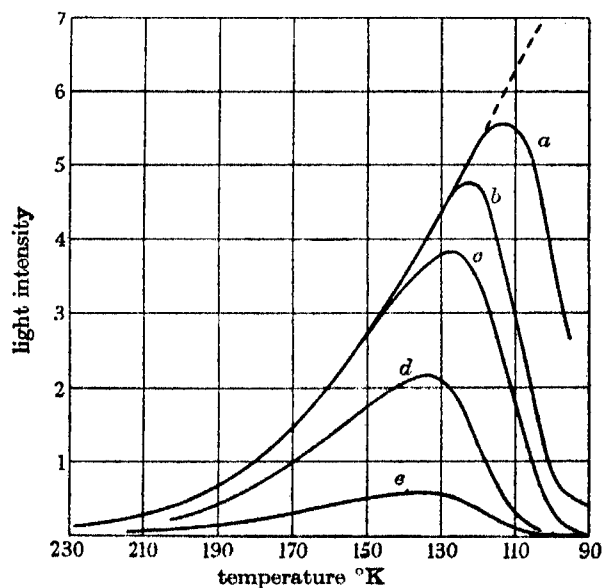


FIGURE 12. Glow curve of SrS after phosphorescence has proceeded for various times at liquid oxygen temperature 90°K . Rate of warming 5° per sec. Dotted curve shows probable form of undecayed peak. [The temperature scale is not quite correct; 10° on the scale correspond in fact to 11° .] Time of decay (a) 5 sec., (b) 25 sec., (c) 140 sec., (d) 10.5 min., (e) 30 min.

The results of figures 10, 11 and 12 are plotted graphically in figure 13 together with data obtained from other experiments. The results agree with theory; the points obtained by decay at 90 and 290° K both lie on

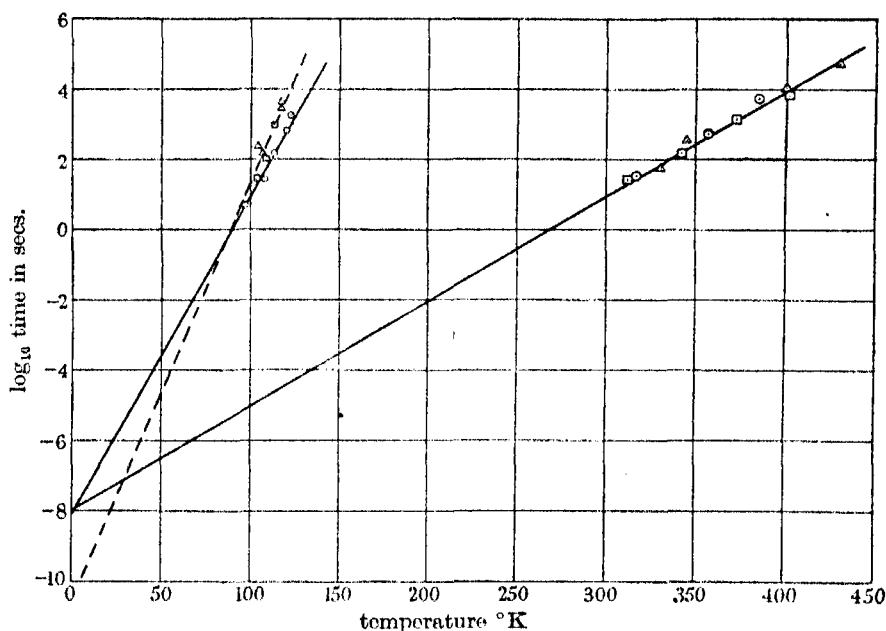


FIGURE 13. Decay of glow curve. The log of the time of phosphorescence decay is plotted against the temperature for which the glow curve has decayed to half its original value. The temperature at which decay took place was about 90 and 290° K. The straight lines drawn through the points cut the log t axis at the value $-\log s$.

Temp. at which decay took place ° K	Phosphor	Glow curves shown in
290	□ ZnS (no. 2)	Fig. 10
293	⊙ SrS (no. 1)	Fig. 11
293	△ SrS (no. 3)	—
90	⊙ SrS (no. 1)	Fig. 12
90	□ Cadmium borate Mn	—
90	△ Cadmium chlorophosphate Mn	—

straight lines which cut the time axis at the same point. s appears to be $\sim 10^{8 \pm 1} \text{ sec.}^{-1}$ for both zinc and strontium sulphides. The results for the cadmium phosphors appear higher ($\sim 10^{11} \text{ sec.}^{-1}$), but the experiments in this case were not made so carefully.

OPTICAL PROPERTIES OF TRAPPED ELECTRONS

Electrons in traps in a phosphor should give rise to an absorption band which will not be present in the unexcited phosphor. Absorption of a quantum of radiation of frequency ν will cause the trapped electron to be ejected from the trap if $h\nu > E$, where E is the depth of the trap. If after absorption of such radiation the phosphor is warmed, the intensity of the glow (which is proportional to the number of trapped electrons) will be reduced below its normal value, provided that the radiation is not of sufficiently short wave-length to re-excite the phosphor. The minimum frequency which will cause this effect provides a value for E , since $h\nu_{\text{min.}} = E$. It is clear, however, from arguments which need not be repeated here (ref. de Boer and van Geel 1935, Mott and Gurney 1940, p. 160), that such a value for E will be greater than the value obtained from experiments on the thermal release of electrons from traps.

Preliminary experiments have been made to find values of E for different temperatures of thermal glow. The method used was a combination of the thermal glow experiment and a photographic method due to Dahms (1904). As expected, E was roughly proportional to the temperature of the glow; this applied for different glow peaks and within single peaks covering a wide temperature range.

Phosphors cooled by liquid air, or liquid hydrogen, might be useful in photographing the infra-red spectrum from 2–10 μ . A sheet of phosphor would be exposed to the infra-red, then placed over a photographic plate and warmed a few degrees.

TABLE 1. OPTICAL AND THERMAL ACTIVATION ENERGIES
FOR A CaSrS-Bi PHOSPHOR

Glow temperature ° K	λ maximum	E (optically) eV	E (thermally) eV
400	1.0 μ	1.2	0.75
250	1.4 μ	0.9	0.5
130	2.0 μ	0.6	0.3

Table 1 shows specimen results for a CaSrS-Bi phosphor. The temperatures shown correspond to the maxima of three large peaks in the glow spectrum. Near the maximum wave-length the effect of the radiation increases rapidly with decrease of wave-length. Several zinc sulphide phosphors gave similar effects. Zinc silicate, zinc mesodisilicate and cadmium silicate and borate were much less sensitive to the radiation.

There was in no case an indication that the effect ceased for short wave-lengths; hence it is reasonable to suppose that the radiation which excites

luminescence in a phosphor is active also (but in general to a lesser extent) in ejecting electrons from traps. Luminescence produced after trapping and ejection from the trap requires the absorption of two quanta to produce one luminescence quantum; hence the efficiency will be lower than usual.

Mott and Gurney (1938) have suggested that the decrease in apparent luminescence efficiency of some phosphors with decreasing temperature may be due to reduction in the absorption coefficient of the phosphor when most of the electrons from the luminescence centres are held in traps. This effect is observed for ZnS-Mn and CaS-Bi and SrS-Bi phosphors. If the explanation is correct, there should be a rapid rise in fluorescence at temperatures associated with well-defined peaks in the thermal glow spectrum; such an effect is, however, not apparent. It may be significant that the phosphors which show decrease in luminescence at low temperatures are those which seem to have empty luminescence centres in the unexcited phosphor (see Paper II).

We wish to thank Professor M. L. Oliphant, F.R.S., for encouragement and provision of facilities for this work. We are also indebted to Professor Peierls for much valuable discussion and help with the theoretical aspects, and to Professor N. F. Mott, F.R.S., for discussion of the physical processes involved. Our thanks are also due to the Warren Research Fund Committee of the Royal Society for a grant to one of us (M. H. F. W.) and to H. C. Cole, who has given skilled assistance in the making of apparatus and much valuable help in the taking of observations.

REFERENCES

- Berg, W. F. 1939 *Trans. Faraday Soc.* **35**, 445.
de Boer, J. H. and van Geel, Ch. 1935 *Physics*, **2**, 286.
Bürger, W. and Flechsig, W. 1931*a* *Z. Phys.* **67**, 42.
——— 1931*b* *Z. Phys.* **69**, 637.
Dahms, A. 1904 *Ann. Phys., Paris*, **13**, 447.
Fonda, G. R. 1939 *J. appl. Physics*, **10**, 408.
Frenkel, J. 1938 *Techn. Phys. U.S.S.R.* **5**, 685.
Herman and Hofstadter 1940 *Phys. Rev.* **57**, 936.
Hilsch, R. 1937 *Proc. Phys. Soc.* **49**, Extra part, 40.
Ives, E. and Luckeish, M. 1911 *Astrophys. J.* **34**, 173.
Johnson, R. P. 1939 *J. opt. Soc. Amer.* **29**, 387.
Lenard, P., Schmidt, F. and Tomaschek, R. 1928 *Handb. Exp. Phys.* **23**.
Mott, N. F. and Gurney, R. W. 1938 *Trans. Faraday Soc.* **35**.
——— 1940 *Electronic Processes in Ionic Crystals*. Oxford: Univ. Press.
Mulder, P. J. 1938 *J. Franklin Inst.* **225**, 527.
Pohl, R. W. 1937 *Proc. Phys. Soc.* **49**, Extra part, 3.
Randall, J. T. and Wilkins, M. H. F. 1945 *Proc. Roy. Soc. A*, **184**, 347.
Seitz, F. 1938 *J. chem. Phys.* **6**, 150.

Phosphorescence and electron traps

II. The interpretation of long-period phosphorescence

By J. T. RANDALL (*Warren Research Fellow*) and M. H. F. WILKINS
Physics Department, University of Birmingham

(Communicated by M. L. Oliphant, F.R.S.—Received 6 November 1940)

This paper is concerned with the measurement and theory of long-period phosphorescence in solids. The delayed emission is, in such cases, due to the time electrons spend in traps or metastable states before returning to the luminescence centres. The theory shows how the observed decay laws are related to the trap distributions found in Paper I. In alkaline earth phosphors there are equal numbers of traps at all depths over the range considered, and the expression developed shows that the intensity of phosphorescence should be inversely proportional to time. For zinc sulphide phosphors an exponential trap distribution has been found, and this leads to a simple inverse power law for the theoretical decay curve. Measurements of decay on these phosphors utilizing an electron multiplier have confirmed the theory, and put on a quantitative basis the connexion between thermoluminescence and long-period decay. At the same time the work provides the first satisfactory and detailed explanation of long-period decays. The extent to which retrapping of electrons modifies the picture given is discussed.

INTRODUCTION

Phosphorescence is of two main types. In the one type the delay in emission of light is due to the time an electron spends in the luminescence centre in an excited state. This kind of phosphorescence lasts only for a few milliseconds, and gives an exponential decay curve which is temperature independent (Randall and Wilkins 1945). In the present work we are concerned with phosphorescence of the other type in which the delay in emission is due to the time an electron spends in a trap. The phosphorescence in this case may continue many hours or days. If the traps are of one depth only, the decay curve will be exponential and highly temperature dependent as in the case of KCl-Tl (Bürger and Flechsig 1931 *a* and *b*); but in most cases many different trap depths are involved, and the resulting phosphorescence emission is given by the summation of the separate phosphorescence emissions (cf. Curie 1939) due to each trap depth. The form of the decay curve and its dependence on temperature then depend on the distribution of trap depths.

In the previous paper, *Phosphorescence and electron traps*, I (to be referred to as Paper I), we have described a method by which the depth of electron traps in a phosphor may be found. In this paper we shall show how the decay curve of phosphorescence is related to the distribution of trap depths. In the theory in this paper and in the previous paper, it is assumed that retrapping of electrons (that is, electrons freed from traps being recaptured by traps before they reach luminescence centres) is a negligible factor, and the extent to which this assumption is justified is discussed in a later section of this paper.

THEORETICAL DERIVATION OF PHOSPHORESCENCE DECAY LAW FOR VARIOUS TRAP DISTRIBUTIONS

(i) *Single trap depth*

If n is the number of electrons in traps of depth E , then the intensity I , of phosphorescence due to traps of depth E , is (from equation (3), Paper I) given by

$$I = dn/dt = -nse^{-E/kT}.$$

Integrating we have

$$I = -n_0 se^{-E/kT} \exp[-ste^{-E/kT}], \quad (1)$$

or

$$I_t = I_0 \exp[-ste^{-E/kT}].$$

This equation gives the intensity of phosphorescence due to a single trap level (cf. Bunger and Flechsig 1931a).

(ii) *Uniform distribution of trap levels*

In most phosphors there are trap levels at many different depths. It is required to find the phosphorescence for a phosphor in which there is an equal number of traps of all depths: that is, the trap distribution is uniform. If N_E is the number of traps between depths E and $E+dE$, then the phosphorescence at time t is given by (from equation (1))

$$I_t = \int_0^\infty N_E se^{-E/kT} \exp[-ste^{-E/kT}] dE,$$

which on integration gives directly

$$I_t = nkT/t(1 - e^{-st}).$$

When $st \gg 1$, that is, after a microsecond, e^{-st} is negligible and

$$I_t = nkT/t. \quad (2)$$

The decay of many phosphors approximates to this law. In simple terms the reciprocal relation may be derived as follows. The intensity of phosphorescence emission due to release of electrons from traps of a certain depth is proportional to the rate of release of electrons from those traps. The rate of release of an electron from a trap is inversely proportional to the mean time an electron spends in the trap. When there is a continuous distribution of trap depths, at time t during the phosphorescence, most of the light emission is due to traps in which electrons spend a mean time t , because the shallower traps are mostly emptied by that time and the deeper traps release electrons too slowly to contribute much to the phosphorescence. If there is an equal number of traps at all depths, the phosphorescence is at all times proportional to $1/t$.

(iii) *Quasi-uniform trap distributions*

For a uniform trap distribution the product It is constant. The previous paragraph may be summarized in the form $It \propto N_t$, where N_t is the number of traps in which electrons spend a mean time t . This relation holds whether N_t is constant or slowly varying.

N_t plotted against E is, however, the trap distribution curve and has approximately the same form as the glow curve. As $N_t \propto It$ and $E \propto \log t$ (equation (1), Paper I), the curve obtained by plotting It against $\log t$ will have the same form as the glow curve. And by equation (6) (Paper I)

$$\log t = \log s \frac{T_G \{1 + f(s, \beta)\} - T}{T},$$

where T_G is the glow temperature and T is the temperature at which phosphorescence takes place. There is, therefore, an almost linear relation depending on s between T_G and $\log t$. If s is chosen suitably, the It - $\log t$ curve may be made to fit the glow curve when glow temperature and time are plotted along the same axis. Such a fitting of the two curves provides a value for the constant s . This link between two independent sets of observations is verified by experiment and is discussed in a later section.

(iv) *Exponential trap distribution*

Many zinc sulphide phosphors have a trap distribution which, in the region we are concerned with, is approximately exponential in form. Thus

$$N_E = A e^{-\alpha E} dE,$$

and as above

$$I_t = \int_0^{\infty} A s \exp[-ste^{-E/kT}] \exp[-E/kT - \alpha E] dE.$$

Putting $\xi = st e^{-E/kT}$, the equation then becomes

$$I_t = (st)^{-\alpha kT} \frac{A s k T}{t} \int_0^{st} e^{-\xi} \xi^{\alpha kT} d\xi.$$

As st is always large ($> 10^9$), the integration may be taken to infinity:

$$I_t \approx (st)^{-\alpha kT} \frac{A s k T}{t} \int_0^\infty e^{-\xi} \xi^{\alpha kT} d\xi,$$

$$I_t \approx f(s, kT) B t^{-(\alpha kT + 1)}. \quad (3)$$

It is clear from equations (2) and (3) that the decay law of phosphorescence for a phosphor with a distribution of trap depths will be little dependent on temperature (W. W. Antonow-Romanowsky 1935; de Groot 1939). While the individual processes involved in phosphorescence are highly temperature dependent, the sum total effect of the processes due to traps of different depths is temperature dependent in quite a different way.

Equation (3) shows that the phosphorescence decays according to a simple inverse power law. For α zero, the case of a uniform trap distribution, the reciprocal law holds; this result agrees with equation (2). When αkT is unity an inverse square law results. In the past such an inverse square law has been correlated with the idea that the excited electrons recombine with the empty luminescence centres in the manner of a bimolecular chemical reaction. The delay in emission of light during phosphorescence is, according to this idea, due to the time taken for an excited electron to find an empty luminescence centre. Calculation shows that the time taken for such a recombination to take place is a small fraction of a second ($< 10^{-5}$ sec.) and for decay times of an order greater than this the time of phosphorescence must be due to the time an electron spends in a trap.

EXPERIMENTAL METHOD OF MEASURING LONG-PERIOD PHOSPHORESCENCE DECAY CURVES

A twelve-stage electron multiplier focused by crossed electric and magnetic fields, and with rubidium surfaces, was used to measure the intensity of phosphorescent light. The current in the last stage of the multiplier never exceeded $50 \mu\text{A}$ and with these conditions the current may be taken as proportional to the light intensity. The current to the galvanometer could be reduced by a known factor so that the galvanometer light spot could be kept on the scale when the phosphorescence was bright. The deflexions were reproducible to within 3% and the sensitivity was so great that for the brightest phosphors measurements could be taken after decay

had proceeded for as long as a week, at which stage the phosphor was almost invisible to a well-accommodated eye.

The phosphor was spread on a glass plate in a layer containing about 0.01 g./sq. cm. Excitation was for over a minute by a 125 W Osira mercury arc in a nickel oxide glass bulb, the lamp being distant 20 cm.

PHOSPHORESCENCE RESULTS

In the first section of this paper relations have been established, for a phosphor, between the trap distribution (or approximately, the glow curve) and the phosphorescence decay curve. Two relations have been established: for quasi-uniform trap distributions, the glow curve will have the same form as the $It\text{-}\log t$ phosphorescence curve, and for the special case of an exponential distribution the decay will follow a simple inverse power law. These relations have been experimentally confirmed.

CALCIUM AND STRONTIUM SULPHIDES

The decay curves for these phosphors are shown in figures 1-4, the product of light intensity and time being plotted against the logarithm of the time. The glow curves are drawn on the same diagrams, a temperature scale being given by equation (6) (Paper I). The value of s was taken as 10^9 sec.^{-1} , but a better fit of two curves might be obtained if s were rather less ($\sim 10^{8.5} \text{ sec.}^{-1}$). In figure 1 the decay and glow curve for a ZnS-Cu phosphor (no. 2) is shown on the same scale as the curve for a CaSrS-Bi phosphor. The correspondence between the glow curves and the decay curves was quite good for figures 1-3. Another specimen of SrS-Bi (results not shown) gave an equally good correlation. In the case of figure 4 the correlation is poor. The phosphor was a CaS-Bi, and the colour of the glow changed strikingly with temperature. On account of this, the form of the glow curve was distorted because the sensitivity of the multiplier altered with wave-length. The correction to the glow curve was obtained by measuring the response of the multiplier to the fluorescence of the phosphor (excited by a constant source of ultra-violet) at different temperatures. The corrected glow curve makes a much better correlation. This procedure was repeated for all the other phosphors including the zinc sulphide discussed in the next section, but in the temperature range considered the correction was found to be small and was not applied. Apart from change in colour of the luminescence the above type of correction may be required because of change in the luminescence efficiency with temperature.

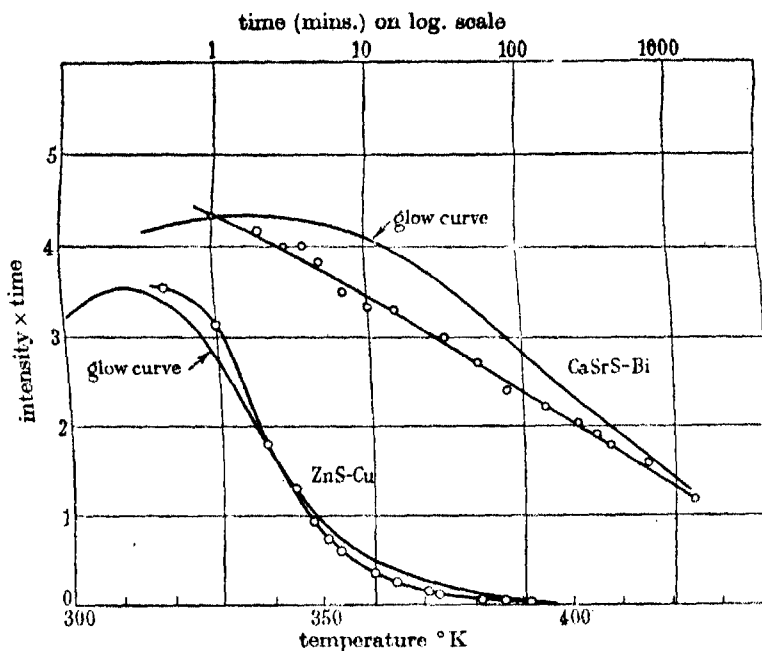


FIGURE 1. Phosphorescence curve and glow curve of ZnS-Cu phosphor (no. 2) and CaSrS-Bi phosphor (no. 4). The correlation between glow curve and phosphorescence curve is clearly shown.

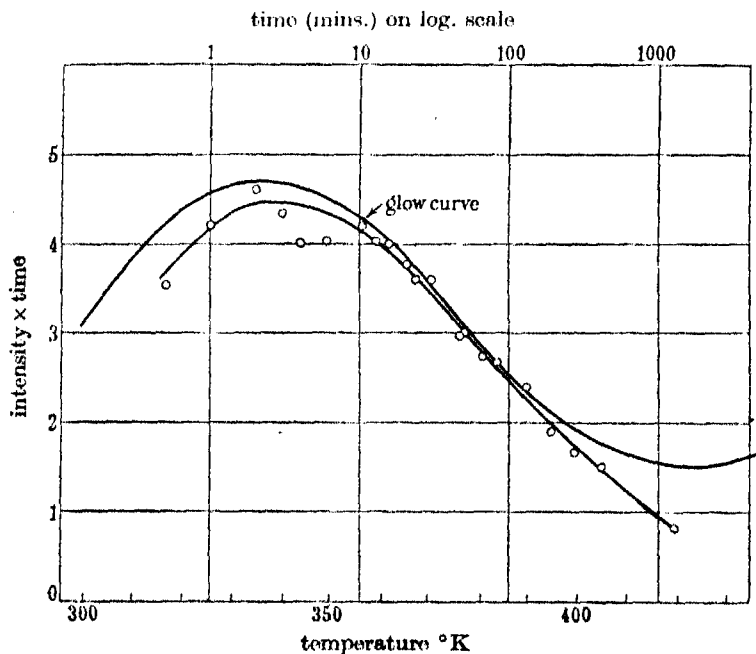


FIGURE 2. Phosphorescence curve and glow curve of SrS-Bi phosphor (no. 1).

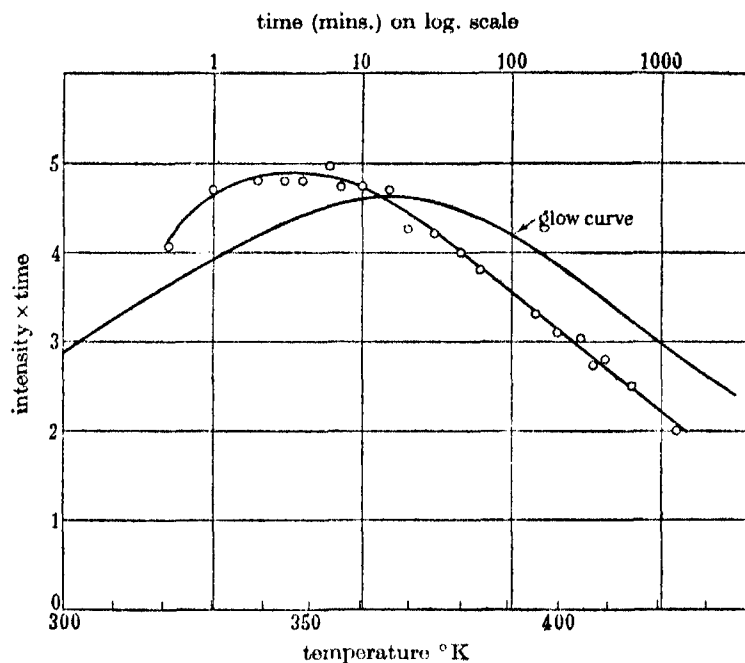


FIGURE 3. Phosphorescence curve and glow curve of CaSrS-Bi phosphor (no. 3).

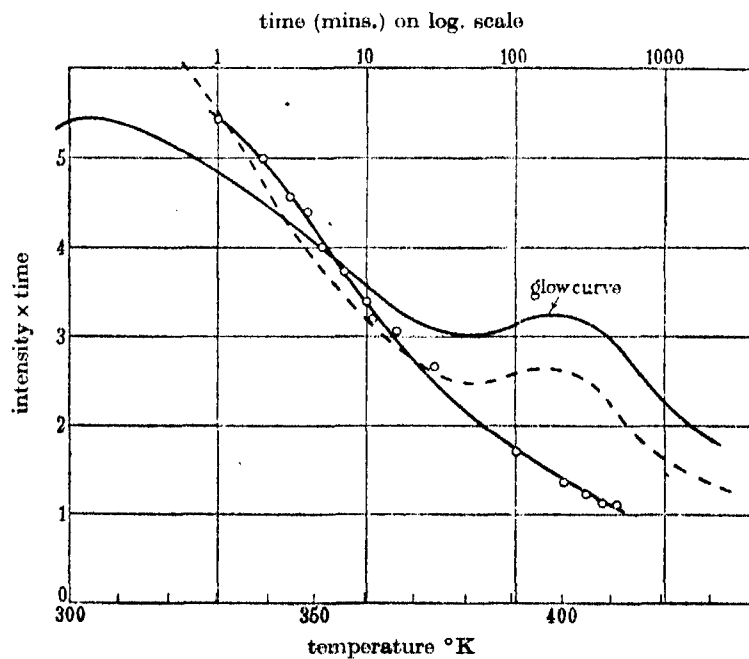


FIGURE 4. Phosphorescence curve and glow curve of CaS-Bi phosphor (no. 6). The dotted line is the glow curve corrected for variation in the colour of the glow.

ZINC SULPHIDE PHOSPHORS

The glow curve for most zinc sulphide phosphors follows an exponential curve in the region which is associated with room temperature phosphorescence, that is, above 300°K (see figure 4, Paper I). As is to be expected, the decay is found to follow a simple power law. In figure 5 the logarithm of the light intensity is plotted against the logarithm of the time: the slope of the line obtained gives the power of the decay law.

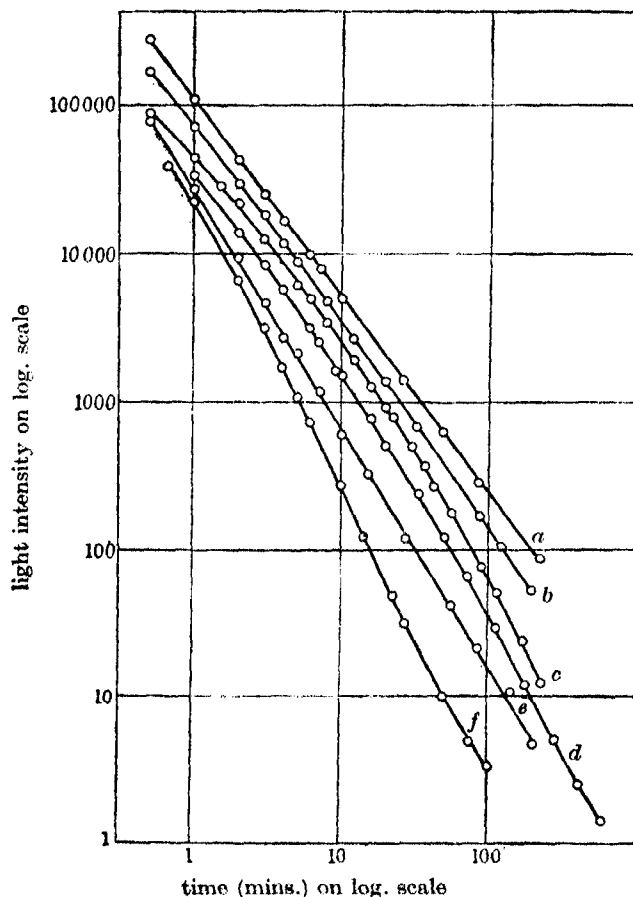


FIGURE 5. Phosphorescence of zinc sulphide phosphors. Various specimens of ZnS-Cu gave rather different decay laws.

- | | |
|--|--|
| (a) ZnS-Cu-Ag at 18°C . | (d) ZnS-Cu at 18°C . |
| (b) ZnS-Cu at 18°C . | (e) ZnS-CdS-Cu at 18°C . |
| (c) ZnS-Cu (no. 2) at 16°C . | (f) ZnS-Cu at 19°C . |

Table 1 shows the observed power law and the value calculated from the glow curve assuming the glow curve to correspond exactly to the trap distribution and taking $s = 10^8 \text{ sec.}^{-1}$. When the glow curve is not exactly exponential the decay is correspondingly not exactly a power law; in such cases the mean value of the theoretical and observed power law is given in the table. The close agreement shown in table 1 should not be over-emphasized, because the form of the glow curve can only be said to approximate to the form of trap distribution and, as is discussed in a later section, retrapping may to some extent cause the decay curve to differ from the form calculated. The different specimens of ZnS-Cu had rather different trap distributions and correspondingly different decay laws.

TABLE 1. DECAY LAWS FOR ZINC SULPHIDE PHOSPHORS

Phosphor	$I \propto t^{-x}$ (value of x obtained from phosphorescence curve)	x calculated from glow curve (taking $s = 10^8 \text{ sec.}^{-1}$)
ZnS-Cu (no. 2)	1.50	1.57
ZnS-Cu	1.38	1.31
ZnS-Cu	1.50	1.51
ZnS-Cu-Ag	1.32	1.30
ZnS-CdS-Cu	1.65	1.57
ZnS-Cu	2.0	2.0

TABLE 2. EFFECT OF INTENSITY OF EXCITATION ON
DECAY OF CaSrS-Bi (NO. 3)

Time of decay in minutes	0.2	1	5	10	20	90
Phosphorescence intensity for exciting intensity				3.0	2.0	1.6	1.5	1.35	1.3
100I divided by phosphorescence intensity for exciting intensity I									

DEPENDENCE OF DECAY LAW ON INTENSITY OF EXCITATION

The results in table 2 show how the decay curve of a CaSrS-Bi phosphor depends on the exciting intensity. For long times of decay the phosphorescence is independent of the exciting intensity provided that the phosphor has been exposed sufficiently long for all the traps to be filled. Under excitation an equilibrium is reached when the rate of release of electrons from traps is equal to the rate of input. For short times of decay shallow traps are involved and, as the rate of release of the electrons is appreciable, these traps are only partly filled when equilibrium is established during excitation. Thus the height and slope of the decay curve increases

with exciting intensity when low exciting intensities or short times of decay are considered. Table 2 shows the experimental results for a CaSrS-Bi phosphor (no. 3).

TOTAL ENERGY STORED IN A PHOSPHOR

Results obtained by Lenard (ref. Lenard, Schmidt and Tomaschek 1928), who measured in absolute units the total thermoluminescence emission from a phosphor, show that there are 4–5 photons emitted per luminescence centre. It is difficult to conceive how more than one photon can be emitted per luminescence centre, and one is led to suppose that Lenard underestimated the number of luminescence centres. We have not, ourselves, made any measurements on this subject, but Harper, Robinson and Bowtell (1940) have measured the decay curve of an alkaline-earth sulphide phosphor in foot-candles and we have made a calculation based on these figures. We find that there is, during phosphorescence at room temperature, about 1 photon emitted per 10^5 – 10^4 Sr atoms in the phosphor. The number of Bi impurity atoms can hardly exceed this figure, so that the number of electron traps in the phosphor of depth greater than 0.2 eV is about equal to or greater than the number of impurity atoms. This result confirms Lenard's figure.

THE CAPTURE OF ELECTRONS BY TRAPS (RETRAPPING)

It has been assumed in the discussion above that an electron released from a trap passes immediately to an empty luminescence centre and that re-trapping does not occur. It is not impossible, however, that an electron might be retrapped several times, and if this were so the theory of phosphorescence we have given above would have to be modified. The extent of retrapping depends on the relative number of empty traps and empty luminescence centres, and the relative probabilities of capture for the two types of centre. Two experimental approaches, which are discussed below, have been useful in studying this question.

(i) *The growth of fluorescence with time*

Consider a phosphor in which all the luminescence centres are filled and all the traps are emptied, and let there be more luminescence centres than traps. The phosphor is exposed to a constant exciting radiation. At the first instant of exposure there are no empty luminescence centres and all the excited electrons will be captured by traps. As the traps are filled, the luminescence centres are emptied and are available for capturing electrons,

so that fluorescence results. The fluorescence intensity increases with time and reaches an equilibrium value when all the traps are filled. We will now derive an expression for the growth of fluorescence with time.

Let n be the number of traps per unit volume,

a the exciting light intensity,

N the number of electrons in the conduction band per unit vol.,

n_1 the number of empty luminescence centres per unit vol.

Electrons enter the conduction band at a rate proportional to a , and leave the conduction band by being captured by empty traps and empty luminescence centres.

Thus
$$dN/dt = a - A_2(n - n_1 + N)N - A_1n_1N,$$

where A_1 and A_2 are constants expressing the relative probability of capture of an electron by an empty trap and empty luminescence centre.

If the time of recombination of free electrons with traps and luminescence centres is small, N is negligible compared with n and n_1 , and after the first fraction of a second N may be taken as slowly variable, then

$$dN/dt = 0 \quad \text{and} \quad N = \frac{a}{A_1n_1 + A_2(n - n_1)}. \quad (4)$$

The rate at which traps are being filled is

$$d(n - n_1 + N)/dt = NA_2(n - n_1 + N),$$

which is approximately equal to the rate at which traps are being filled,

$$dn_1/dt = -NA_2(n - n_1). \quad (5)$$

Combining equations (4) and (5)

$$dn_1/dt = aA_2(n - n_1)/A_1n_1 + A_2(n - n_1),$$

and integrating

$$(1 - A_1/A_2)n_1 - A_1/A_2 \cdot n \log(n - n_1/n) = at. \quad (6)$$

This equation expresses the variation of n_1 with t .

The fluorescence emission I is A_1Nn_1 , or using equation (4)

$$I = \frac{a}{1 - A_2/A_1 \cdot (1 - n/n_1)}. \quad (7)$$

To see the form of the I - t curve it is simplest to consider equations (6) and (7) separately.

When n_1 is small, (6) gives $n_1 = at$; therefore

$$I = A_1Nn_1 = A_1/A_2 \cdot a^2t/n. \quad (8)$$

Hence the I - t curve begins as a straight line with slope $A_1/A_2 \cdot a^2/n$. When n_1 approaches n ,

$$(n - n_1) \propto e^{-aA_2/A_1 n}, \quad (9)$$

which means that near the equilibrium value the fluorescence intensity follows an exponential curve. Both equations (8) and (9) provide simple ways of finding A_1/A_2 .

Factors which have not been taken into account in the theory are: non-uniform illumination of the phosphor grains, and the finite rate of escape of electrons from traps due to thermal motion and due to absorption of the exciting light.

(ii) Experimental results

A ZnS-Cu phosphor (no. 2) was heated in the dark on the glow apparatus to empty all the traps. Then at room temperature the exciting radiation was switched on. A mercury arc was used with filters of nickel oxide glass and cupric chloride in water, and a Wratten 2a filter was placed over the multiplier which measured the fluorescence intensity.

The curves of figure 6 agree well with the theory given above. The amount of energy stored in the phosphor is given by the area between the rising fluorescence curve and the line drawn at the level of equilibrium fluorescence. This area should be equal to the area under the phosphorescence and glow curves. The amount of stored energy increases with exciting intensity because the electrons have an appreciable rate of escape from the shallower traps, and hence the extent to which these traps are filled increases with the exciting intensity. Table 3 shows the results of figure 6.

TABLE 3. ENERGY STORED IN A ZnS-Cu PHOSPHOR (NO. 2)

Exciting intensity in arbitrary units	$\int Idt$ for glow and phosphorescence	$\int (I_\infty - I_t) dt$ for fluorescence
1	100	98
3	120	110
9	160	148

A_1/A_2 may be calculated from the curve using equations (8) and (9). (a , I , t and n are all measured from the curve in units of length and area.)

The initial part of the curve gives $A_1/A_2 = 1.3$, and the part of the curve approaching equilibrium gives a value 2.2. The two values agree reasonably well. The meaning of the result is that the cross-sections for electron capture of the traps and luminescence centres are about equal.

Figure 7 shows some results for ZnS-Mn at 90° K. The traps in this phosphor are of a depth such that at room temperature very few can be

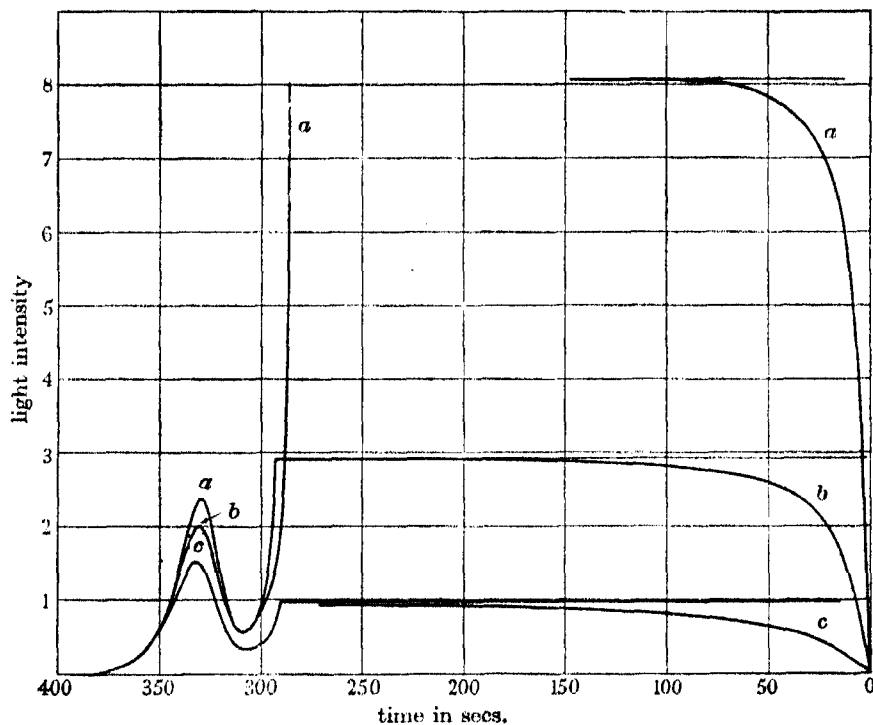


FIGURE 6. Curves for growth of fluorescence, with time for a ZnS-Cu phosphor (no. 2) at 290° K. Intensity of exciting light (a) 9I, (b) 3I, (c) I. The time of excitation was (a) 150 sec., (b) 300 sec., (c) 800 sec. After excitation the phosphorescence and glow curve were recorded.

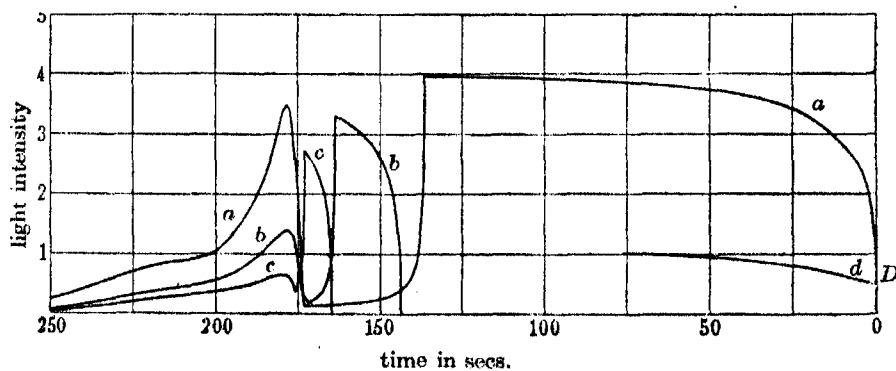


FIGURE 7. Curves for growth of fluorescence with time for ZnS-Mn at 90° K. Different times of excitation with the same exciting intensity were used for curves a, b and c. After the exciting light has been out off the phosphorescence and glow curve has been recorded. Curve d is for one-third the exciting intensity and in this case the discontinuity in the curve is shown clearly at D.

filled, and at room temperature the fluorescence rises to 95 % of its equilibrium value in less than a second. At 90° K most of the electrons are stable. In this case the fluorescence appears to rise instantaneously from zero to about 40 % of the equilibrium fluorescence value. As the galvanometer has a period of 2 sec., the nature of this sudden rise could not be followed. After the sudden rise the fluorescence increases slowly and tends exponentially to an equilibrium value. Table 4 shows the values of the stored energy after different times and intensities of excitation. In all cases $\int I dt$ for the glow curve is twice $\int (I_{\infty} - I_t) dt$ for fluorescence. Measurements on the brightness of fluorescence at different temperatures indicate that the luminescence efficiency changes with temperature. If a correction is made on this basis, the above factor of 2 is reduced to 1.5, but the expected equality is not obtained. A possible explanation is that the luminescence efficiency is low during fluorescence because some of the fluorescence is produced by ejecting electrons from traps.

To explain the initial instantaneous rise of fluorescence, two explanations suggest themselves. The first explanation is that some of the electrons when excited may reach excitation levels in the luminescence centre and return to the ground state without passing into the conduction band, and these electrons would then give rise to an almost instantaneous fluorescence. Alternatively, there may be empty luminescence centres in the crystal before excitation begins, thus allowing electrons ejected into the conduction band to return almost immediately to the ground state. This idea can also explain the form of the glow curve when the traps are only partly filled (see figure 7 and below).

TABLE 4. ENERGY STORED IN ZnS-Mn AT 90° K

Exciting intensity in arbitrary units	$\int I dt$ for glow and phosphorescence	$\int (I_{\infty} - I_t) dt$ for fluorescence
3	400	175
3	150	75
3	80	47
1	18	10
1	38	20

(iii) *The glow curve when the traps are partly filled*

There are two stages in producing a glow curve when the time of excitation is not sufficient for all the traps to be filled. First, during excitation electrons are captured by traps, the number of electrons caught in traps of a certain

depth being proportional to the number of such traps and the trap cross-section for electron capture. Secondly, the phosphor is warmed and some of the electrons liberated are immediately captured by luminescence centres and a proportion are captured by deeper traps. The result of this retrapping is that the partly charged up glow curve is relatively higher at high temperatures than it would have been if no retrapping had taken place.

Examination of partly charged up glow curves of the type shown in figure 7 indicates that the cross-section of all traps of different depth is about the same, and, to account for the small amount of retrapping apparent, it is

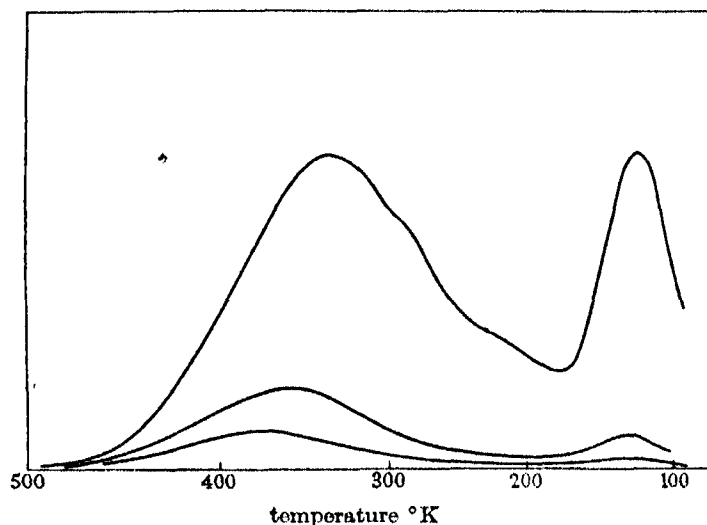


FIGURE 8. Glow curve for a CaSrS phosphor (no. 4). The two lower curves correspond to partial filling of the traps and the top curve corresponds to saturation conditions. The relative heights of the two peaks in the top curve is clearly different from the relative heights in the lower curves.

necessary to assume that there are empty luminescence centres in the un-excited phosphor—there being about three traps to each empty luminescence centre. Similar results are also indicated for CaS and SrS phosphors (see figure 8), which also give an instantaneous rise of fluorescence. These conclusions must, however, be regarded as somewhat tentative, as further study is required.

Figure 9 shows the fluorescence intensity time curves for an SrS and CaSrS phosphor. The values of A_1/A_2 given by the curve are 1.5 and 3. The specimens were heated to about 700° K in the dark before the experiment, to ensure that all the traps were emptied.

The fact that the cross-section of traps and luminescence centres is about the same suggests that the process of capture is similar in both cases. The capture into a trap is almost certainly non-radiative (though traps might fluoresce in the infra-red at sufficiently low temperatures); hence it is to be expected that the capture of an electron by a luminescence centre is also non-radiative. The only way this can occur is for the electron in the conduction band to be first captured into an excited state of the luminescence centre and then to make a luminescence transition within the centre from the excited state to the ground state (Mott and Gurney 1940, p. 108). Apart

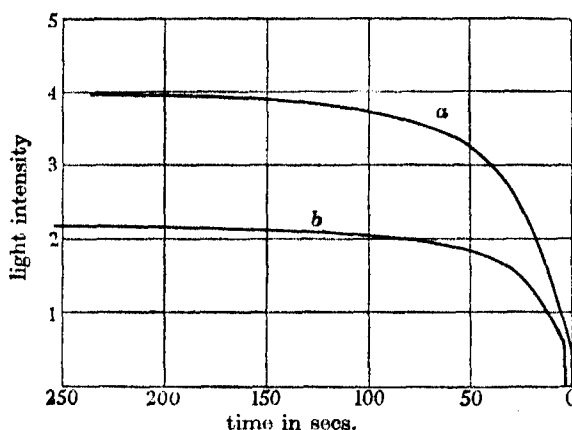


FIGURE 9. Growth of fluorescence with time for CaSrS-Bi phosphor. Curve *a*, phosphor no. 4; curve *b*, phosphor no. 3.

from the experimental evidence discussed here, this process seems more likely than a direct radiative transition from the conduction band to the ground state of an impurity atom. Support for this idea is also to be found in the fact that phosphors activated with manganese photoconduct and give a quick exponential phosphorescence (Randall and Wilkins 1940). The exponential decay is associated with the lifetime of an electron in the excited state of a manganese ion.

THE EFFECT OF RETRAPPING ON THE PHOSPHORESCENCE DECAY CURVE

Figure 11, curve (*d*) (Paper I) shows the amount of stored energy which is liberated from a SrS-Bi phosphor (no. 1), after phosphorescence has proceeded for 80 min. The area under that glow curve is proportional to the number of empty luminescence centres. The area between curve (*d*) and the curve (*a*) from 370° K upwards is proportional to the number of empty traps

in which a trapped electron has a mean lifetime of more than 30 min. The ratio of these areas is 4:1. Similarly considering the decay after 10 min., the ratio of the number of empty luminescence centres to the number of empty traps of lifetime greater than 3 min. is 3:1. The value of A_1/A_2 is about 1. Thus at time t during phosphorescence about 75 % of the electrons being liberated from traps pass almost immediately to luminescence centres, and 25 % are delayed by retrapping for a mean time of about $\frac{1}{3}t$. Hence in this case, which is typical of CaS and SrS phosphors, retrapping has very little influence on the decay curve. There will be much less influence still if (as seems likely) there are more empty luminescence centres than there are trapped electrons.

The ZnS-Cu phosphor (no. 2) contains an equal number of trapped electrons and empty luminescence centres. In this case A_1/A_2 is again about 1, and using the data in figure 10 (Paper I) it has been calculated that at various times t of decay almost 50 % of the electrons being liberated from traps will be retrapped into traps of mean lifetime greater than $\frac{1}{3}t$. 50 % of the electrons give phosphorescence emission after a mean delay due to retrapping of about $\frac{1}{3}t$. If the phosphorescence is calculated from the trap distribution, it may be said that the emission which would take place at time t , if there were no retrapping, does in fact take place at time $t + \frac{1}{3}t$. It is rather surprising that, even in this case, retrapping does not appear to have a very marked effect on the form of the decay curve.

MAGNITUDE OF THE CONSTANT s

It has been said in Paper I that s should be of the order rather less than that of the thermal vibration frequency of a crystal ($\sim 10^{12}$ sec.⁻¹). Theory cannot, however, give anything more than a rough guess for the magnitude of s until a particular model of a trap is considered. The most accurate experimental value of s , $10^{9.5}$ sec.⁻¹, has been obtained from measurements of the phosphorescence of potassium chloride (Bünger and Flechsig 1931*a*). Our measurements give a value of about $10^{8 \pm 1}$ sec.⁻¹ for many specimens of zinc sulphide and calcium and strontium sulphides, measurements having been made at 290 and 90° K. We have deduced our value of s from the experimental material by approximate methods, and for this reason the accuracy of the value we give must not be stressed too much, although two methods, phosphorescence measurements and decay of glow peak measurement, give the same value of s .

A rough value for s has been obtained for F' centres in alkali halides (Mott and Gurney 1940, p. 136) from the decrease of photoconductivity at

low temperatures. The value obtained is $10^{10} \text{ sec.}^{-1}$, but several assumptions are made in the calculation. Another value for the same s has been obtained from measurements of the movements of photoelectrons in alkali halides. The cross-section of the F centre trap is calculated from the distance which a photo-electron travels before it is captured by a trap. The s value is related to the cross-section by a formula given by statistical mechanics (Mott and Gurney 1940, p. 108). The formula cannot, however, be taken as exact because the interaction of the trapped electron with the lattice is not taken into account in deriving the formula. The cross-section of the F centre is 10^{-18} cm^2 and by the formula s is $10^{11} \text{ sec.}^{-1}$. If s is 10^8 sec.^{-1} in a phosphor, the cross-section of a trap would be 10^{-18} cm^2 . We have shown above that the cross-sections of traps and luminescence centres are about the same in phosphors. If the cross-section of a luminescence centre is as small as 10^{-18} cm^2 , the bimolecular theory of short-period phosphorescence may be true. The cross-section of a trap given by measurements of the photoelectric current may correspond to capture of an electron into an excited state of a centre and the cross-section for capture into the ground state may be smaller because an electron may be captured into an excited state and released again several times before it falls to the ground state of the centre.

Our thanks are due to Professor M. L. Oliphant, F.R.S. for his provision of facilities for this work and for his encouragement. We are greatly indebted to Professor R. Peierls, who has given much time to the discussion of theoretical aspects and has helped with numerous mathematical points. Professor N. F. Mott, F.R.S. has also provided us with some valuable suggestions. H. C. Cole has again given us his skilled assistance. Thanks are also due to the Warren Research Fund Committee of the Royal Society for a grant to one of us (M. H. F. W.).

REFERENCES

- Antonow-Romanowsky, W. W. 1935 *Phys. Z. Sowjet.* **7**, 336.
 Büniger, W. and Flechsig, W. 1931a *Z. Phys.* **67**, 42.
 ——— 1931b *Z. Phys.* **69**, 637.
 Curie, M. 1939 *Trans. Faraday Soc.* **35**, 114.
 de Groot, W. 1939 *Physica*, **6**, 275.
 Harper, W. E., Robinson, M. B. and Bowtell, J. N. 1940 *Trans. Illum. Engng Soc.* **5**, 57.
 Mott, N. F. and Gurney, R. W. 1940 *Electronic processes in Ionic Crystals*. Oxford Univ. Press.
 Randall, J. T. and Wilkins, M. H. F. 1945 *Proc. Roy. Soc. A*, **184**, 347.

Short period phosphorescence and electron traps

BY G. F. J. GARLICK AND M. H. F. WILKINS

Department of Physics, The University, Birmingham

(Communicated by N. F. Mott, F.R.S.—Received 4 August 1942—

Revised 13 May 1943)

The decay of phosphorescence of various types is measured during the first few milliseconds of the process. Phosphorescence with exponential decay occurs when luminescence is due to an optical transition of a type normally forbidden. Experimental proof of this is provided by a study of the ruby. A further study is made of the effect of temperature on exponential decays. Some phosphors give complex decays in which a temperature-dependent decay process due to thermally metastable states or traps is superimposed on the exponential decay. An experimental separation of these decay processes is described for the case of ZnS-Mn.

It had been believed that in hyperbolic decays, characteristic of ZnS-Cu and ZnS-Ag, the time constant of the process was due to the time that an excited electron spent in moving through the phosphor. The experiments described here show, however, that this type of decay is due primarily to the time electrons spend in traps; therefore the bimolecular theory of phosphorescence is largely rejected. The main experimental results show the temperature dependence of the decay curve and the correlation between the decay curve and the amount of light energy which may be stored in traps in the phosphor at low temperatures. An explanation is given of the change of shape of the decay curve with temperature.

INTRODUCTION

When light falls on a body some of its energy may be absorbed and re-emitted as light of a longer wave-length. This phenomenon is called luminescence. In some solid materials there is a time lag between the absorption of the exciting energy and the resulting luminescence, and luminescence may continue after the exciting light is removed. Such delayed luminescence is called phosphorescence and solids which show it are called phosphors. The present work is concerned with the elucidation of the processes which produce phosphorescence in solids.

Becquerel (1867) began the study of phosphorescence processes by measuring how the intensity of phosphorescence after-glow decreased with time after excitation had ceased. He found that phosphorescence decay curves so obtained were of two types, exponential and hyperbolic. Exponential decay occurs when the radiating electron does not leave the excited atom, and may occur if the atoms are isolated from one another as in a gas or

aggregated in a solid. More interest lies in the interpretation of hyperbolic decays which may be of very great duration and which are peculiar to the solid state. Becquerel suggested the bimolecular theory of phosphorescence which, stated in modern terms, attributes the delay in emission of phosphorescence to the time an excited electron spends in moving through the phosphor from one atom to another. Lenard, Schmidt & Tomaschek (1932) studied alkaline earth sulphide phosphors and showed qualitatively that the phosphorescence which lasted minutes and hours was due to slow release of excited electrons from thermally metastable levels. However, many later investigators of phosphorescence (e.g. Lewschin & Antonow-Romanowsky 1934) attempted to explain their results in terms of the bimolecular theory. Recently, Randall & Wilkins (1945 *b, c*) made a more complete study of the thermally metastable levels (now called electron traps) in phosphors and gave a quantitative explanation of long-period phosphorescence and showed that the experimental results, e.g. Lewschin & Antonow-Romanowsky (1934), which had provided the main support for the bimolecular theory could be more adequately explained in terms of electron traps. Further work (not yet published) in this laboratory on the temperature dependence of long-period phosphorescence decay curves has provided more proof of the correctness of these ideas. It was, however, still an open question whether the time of movement of an electron through the phosphor was a major factor in short-period phosphorescence lasting a few milliseconds. De Groot (1939*a*), studying such phosphorescence, had explained his results in terms of the bimolecular theory but showed that electron traps could have a large effect on the form of the decay. The present work shows that the ideas of Randall & Wilkins (1945 *b, c*) which explain long-period phosphorescence in terms of electron traps do also, to a large extent, explain short-period phosphorescence, and we find practically no evidence that electrons take appreciable time in moving about a phosphor and thus give rise to a bimolecular decay. Our experimental results illustrate the properties of this type of trap decay and, partly for comparison, results are given for exponential decays: complex decays due to the combined action of trap and exponential mechanisms are also shown to occur.

The study of phosphorescence decay processes is of general interest in that it provides information about electronic energy levels and movements of electrons in solid insulators. The concept of electron traps, which we use to explain phosphorescence, is also used to explain the phenomena of photoconductivity, photochemical change, and dielectric breakdown in solids. The reality of electron traps has been made specially clear by the work of Pohl (1937) on colour centres in alkali halides.

THEORETICAL

To make clear the import of the experimental results which follow later we will first give a semi-theoretical discussion of phosphorescence. Parts of this discussion have appeared before (Randall & Wilkins 1945) but are included for completeness. The generally accepted idea of luminescence processes is based on the following model. In figure 1 the ordinate is energy of an electron and the abscissa denotes the position of the electron in an insulating solid. Level *A* is the ground state of an electron in an ion in the solid. When light is absorbed by the ion the electron may be raised in energy to *B* an excited state of the ion. When the electron returns from *B* to *A* a

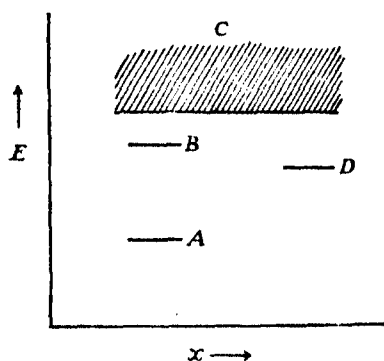


FIGURE 1.

luminescence quantum is emitted. The ion, or group of ions, which contains the levels *A* and *B* and in which the luminescence transition takes place is called the luminescence centre. In many phosphors the luminescence centre is an impurity ion. If a large quantum of energy is absorbed the electron may be raised from *A* into the region *C*, called the conduction band. In this state the electron is free from the luminescence centre and may move through the solid which can then, to a slight extent, conduct electricity and is said to be photoconducting. An electron in the conduction band may fall into a trap level *D* and cannot escape until it is supplied with a quantum sufficient to raise it into *C*. Once in *C* the electron may be recaptured by a trap or return to *B* and then produce luminescence. Phosphorescence is produced by the slow release of electrons from *D* by heat quanta supplied from the thermal vibrations of the phosphor.

We will now consider the type of phosphorescence decay to be expected if the excited electron spends time in the excited state *B*, in the trap state *D*, or in the conduction band *C*.

(1) *Exponential decay due to the time an electron spends in an excited state*

As we have said above, the excited electron may be raised from one energy level in an atom or ion to a higher level in the same ion. The phosphorescence emission I will be proportional to the rate at which excited electrons return to the ground state. Hence, if n is the number of excited electrons and p is the probability of an excited electron returning to the ground state,

$$I = -dn/dt = pn, \quad \text{and hence} \quad I_t = I_0 e^{-pt},$$

where I_0 and I_t are the intensities of phosphorescence at the time when excitation ceases and at time t afterwards. The decay curve is therefore exponential. For ordinary optical transitions, p is large and the phosphorescence will disappear in $\sim 10^{-8}$ sec. (e.g. many luminescent gases), but if the transition from the excited state to the ground state is of a type normally forbidden, p may be relatively smaller and the phosphorescence will be of greater duration.

We may distinguish this type of decay by the following properties which are illustrated in the experimental section and figures 3-6:

- (1) The decay curve is exponential.
- (2) The form of the curve is determined by p and hence will be independent of the conditions of excitation such as intensity, wave-length and duration of excitation.
- (3) The decay curve is independent of temperature if, as may often be the case, the levels in the luminescent ion are unaffected by the thermal vibrations of the solid.
- (4) The phosphor may be non-photoconducting as there is no need for an electron to leave the luminescent ion during the phosphorescence process.
- (5) If the phosphor photoconducts, an additional decay process will be present due to trapping of electrons which escape from the luminescent ions.
- (6) The decay is characteristic of a given ion in the luminescent solid and may be largely independent of the solid in which the ion occurs.

(2) *Phosphorescence decay due to time an electron spends in traps*

If the time an excited electron spends in the conduction band and in the excited state of the luminescence centre is negligible, phosphorescence will be due entirely to the time an electron spends in traps. In a phosphor containing traps, light energy may be stored indefinitely if the temperature is kept low, but if sufficient thermal energy is available the trapped electrons will escape and produce phosphorescence. In general, the decay due to traps is a somewhat complicated phenomenon, the shape of the decay curve depending on the mode of excitation and the extent to which the traps are

filled. The distinguishing characteristics of this kind of decay curve are as follows:

- (1) The decay curve will vary with temperature.
- (2) The decay may be affected by absorption of infra-red light which frees trapped electrons.
- (3) The decay curve resulting from prolonged excitation will vary with the intensity and wave-length of the exciting light.
- (4) The shape of the decay curve will vary with the duration of excitation.
- (5) The phosphor will photoconduct as the excited electron travels through the phosphor from luminescence centre to trap.

We will discuss several of these characteristics in more detail below.

Variation of decay with intensity of excitation

The probability p of an electron escaping from a trap is given by

$$p = se^{-E/kT}, \quad (1)$$

where the activation energy E is called the trap depth and is the minimum amount of energy which may remove the electron from the trap, k is Boltzmann's constant, T is the absolute temperature and S is a constant $\sim 10^{-8} \text{ sec.}^{-1}$ (Randall & Wilkins 1945*b*). Most phosphors contain traps with depths distributed over a wide range, and the form of this distribution will determine the shape of the decay curve. One may calculate long after-glow decay curves by assuming that, after prolonged excitation, all the electron traps are filled (Randall & Wilkins 1945*c*). But for short after-glows the traps are only partly filled even during prolonged excitation because equilibrium will be established when the rate of supply of electrons to traps is equal to the rate of escape and the extent to which traps are filled will therefore depend on their depth E . Traps with large E will become almost completely filled because the rate of escape of electrons from these deep traps is small. Shallower traps will be partly filled because the rate of escape of electrons from these traps is appreciable. When E is very small the probability of escape from a trap is large and most of the traps are empty. When the exciting light is cut off, the electrons continue to escape from the traps and produce the after-glow. The shape of the decay curve after prolonged excitation will depend on the intensity of excitation. If the exciting intensity is increased, more of the shallow traps will be filled, the deep traps being already almost filled to capacity; hence, during the after-glow, an excited electron will spend on the average a shorter time in traps and the after-glow will decay more rapidly (see Johnson 1939), although the intensity of phosphorescence will be greater.

Variation of decay curve with time of excitation

Let us consider qualitatively the type of decay which results from a periodic pulse excitation as distinct from prolonged excitation. Let there be equal numbers of traps of all depths in the phosphor. The deep traps will be completely filled and will supply a slow component to the decay curve. The shallow traps will be empty before excitation and the pulse will fill shallow traps of all depths to an equal extent if the cross-section for electron capture of the traps does not vary much with their depth. Equal numbers of electrons in traps of all depths produce a $1/t$ decay equation (Randall & Wilkins 1945c). Hence one might expect the initial stages of decay to consist of a $1/t$ decay due to the shallower traps. After excitation of long duration all traps are filled to an equilibrium extent, the shallow traps being filled less than the deeper traps. Pulse excitation fills all shallow traps equally, and hence the very shallow traps are filled more by pulse excitation than by prolonged excitation of the same intensity. Therefore the initial decay due to pulse excitation should be more rapid than the decay due to prolonged excitation of the same intensity.

A bimolecular decay curve is a function of I_0 , see equation 6, and does not vary with the time of excitation. A pulse of high exciting intensity should give the same decay curve as a prolonged excitation by a low intensity if I_0 is the same in the two cases. Hence a bimolecular decay may be distinguished from a trap decay which becomes faster rather than slower when the time of excitation is reduced.

Variation of the decay curve with temperature

The variation of the decay curve with temperature is a complicated phenomenon; in some cases the decay curve does not alter appreciably over a wide temperature range, and in other cases the curve may alter very rapidly with temperature.

It is clear from the discussion above that if the phosphor is exposed to prolonged excitation and then allowed to decay the intensity I_t of phosphorescence at time t will increase as the number of traps in which an electron spends a mean time t is increased. The depth E which a trap requires to possess in order that an electron may be trapped in it for time t varies with the temperature. In most phosphors the number of traps varies with the depth of the trap; hence the number of traps in which an electron spends a mean time t will vary with temperature, and therefore I_t (in practice we measure I_t/I_0) will vary with temperature.

The way in which the number of traps varies with the trap depth may be found by measuring the glow curve (Randall & Wilkins 1945b) which is a

curve of thermoluminescence intensity against temperature, obtained when a phosphor is warmed at a constant rate in the dark after having been exposed at a low temperature to excitation. The number of traps N_E of depth E is proportional to the height of the glow curve, and E is proportional to T_G the glow temperature. This may be seen as follows. When the phosphor is warmed at a constant rate, electrons escaping from traps of depth E cause a peak of light emission which reaches a maximum at a temperature T_G given by

$$t_1 = \frac{1}{S} \exp\left(-\frac{E}{kT_G}\right) \quad (2)$$

(see Randall & Wilkins 1945*b*), where t_1 depends on the warming rate and is ~ 10 sec. for the rate of 2°C/sec. used in our experiments. Therefore the glow curve at temperature T_G is produced in the main by electrons from traps of depth E , and the number of these traps is proportional to the height of the glow curve at that temperature.

To connect the decay curve with the glow curve we combine equations (1) and (2) and find that at temperature T the lifetime t of an electron in a trap of depth E corresponding to a temperature T_G on the glow curve is given by the relation

$$\frac{T_G}{T} = \frac{\log St}{\log St_1}. \quad (3)$$

And, at temperature T , the number of traps in which an electron has a lifetime t is proportional to the height of the glow curve at temperature T_G . Hence, if we measure decay curves at different temperatures and plot I_t/I_0 against temperature the curve obtained will have the same form as the glow curve. Corresponding temperatures on the two curves will be in constant ratio given by (3) and hence, if plotted on the same logarithmic scale of temperature, the one curve will be displaced relative to the other a constant distance along the temperature scale. If the glow curve does not vary much in height over a wide temperature range the decay curve will also not alter over a corresponding range of temperature. The extent to which the decay curve varies with temperature therefore depends entirely on the shape of the glow curve.

We will now consider how the physical processes which take place during the decay will vary with temperature, and in this way we can find how the shape of the decay curve will vary with temperature. For simplicity we will take the case of a phosphor with a glow curve concentrated in one peak (see figure 2). The phosphor is excited periodically by a prolonged pulse of light of constant intensity. At temperature T_a (figure 2), well below the

glow peak, all the traps in the phosphor will be permanently filled by electrons which cannot escape and, as these electrons have been removed from luminescence centres, there will be an equal number of permanently empty luminescence centres. When exciting light raises an electron into the conduction band there will be no empty traps to delay the electron which will, therefore, very quickly be captured by one of the many empty luminescence centres and produce luminescence. As a result, the rise and decay of luminescence will be rapid. At a higher temperature T_0 the rate at which electrons

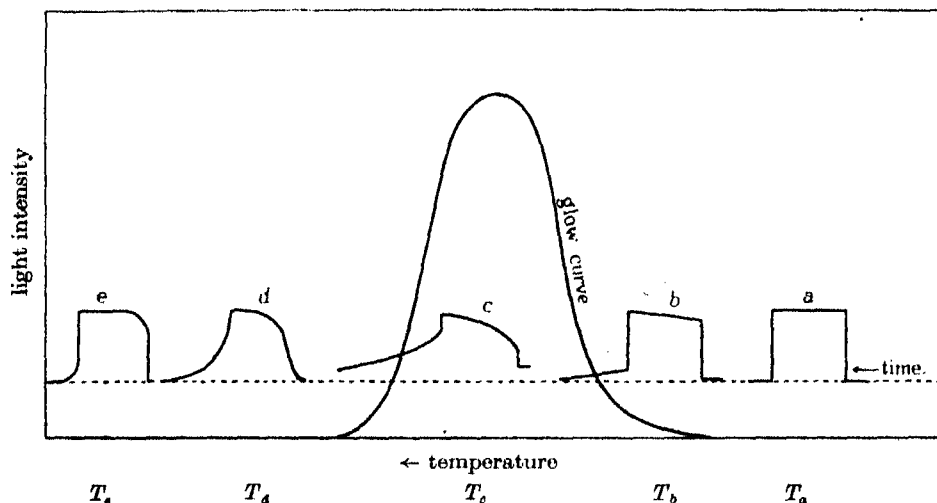


FIGURE 2. Variation of decay curve with temperature. These curves are similar to curves for variation of photocurrent with time at various temperatures in NaCl (Pohl 1937).

escape from the shallower traps becomes appreciable, and this process will produce a faint after-glow which decays slowly. There are, therefore, two distinct parts of the decay curve, a rapid and a slow component. The ratio of these components will alter with temperature. For instance, at a higher temperature T_c , an average fraction r of the traps are filled, hence r luminescence centres are empty and $1 - r$ traps are empty. If the cross-sections for electron capture of an empty trap and of an empty luminescence centre are about equal, r of the electrons raised into the conduction band per unit time during excitation will be captured directly by luminescence centres and $1 - r$ will be captured by traps. Hence, when the excitation ceases, a fraction r of the luminescence will disappear almost immediately and the remainder of the luminescence will disappear as a trap decay. Also, owing to the fact that

a large fraction of the luminescence rises and falls slowly with time, there is a reduction in the extent to which the luminescence varies in intensity during the cycle of excitation and decay. At a higher temperature T_d the phosphorescence decays to zero intensity between the periods of excitation and all the traps are emptied during the period of decay. Hence, at the beginning of each period of excitation there will, at first, be no empty luminescence centres but there will be a large number of empty traps; as a result, an excited electron may be captured by several traps in succession before reaching a luminescence centre. As we see below, in the discussion of bimolecular decay processes, the decay will, in these conditions, possess bimolecular characteristics, the main characteristic being an inflexion on the rise curve of luminescence. At higher temperatures such as T_c the decay curve will become rapidly faster.

(3) *Bimolecular decay produced by an excited electron spending time in moving through the phosphor*

Let us calculate the phosphorescence decay curve produced when an electron spends time in moving through the phosphor assuming that the time spent in traps and in the excited state of the luminescence centre is negligible. If there are no empty luminescence centres or free electrons in the unexcited phosphor, the number n of empty luminescence centres will be equal to the number of excited electrons. The rate of recombination of electrons and luminescence centres is equal to the intensity of luminescence I and is given by

$$I = Avn^2, \quad (4)$$

and during the after-glow

$$I = -\frac{dn}{dt} = Avn^2, \quad (5)$$

where v is the velocity of a free electron and A is the cross-section for electron capture of a luminescence centre. Integration of equation (5) gives the equation of the phosphorescence decay,

$$I_t = \frac{I_0}{(1 + t(I_0 A v))^2}. \quad (6)$$

This formula expresses the rate at which a bimolecular chemical reaction takes place, and hence this theory of phosphorescence is called the bimolecular theory.

When the phosphor is exposed to a constant intensity of exciting light the luminescence intensity increases with time. Initially n is proportional to t , the time of excitation, and hence by equation (4) $I \propto t^2$. The curve of I

against t therefore begins with an upward curvature, but I must tend to an equilibrium value as excitation continues and hence the curve passes through an inflexion.

The distinguishing characteristics of a bimolecular phosphorescence are as follows:

(1) There will be an inflexion on the curve of the rise of luminescence with time.

(2) As indicated by equation (6) the decay curve is independent of the time of excitation and, at a given temperature, is determined by I_0 .

(3) The rapidity of the decay will increase steadily with temperature, for v will be proportional to the square root of the absolute temperature.

In many cases one would not expect the bimolecular theory to be true. First, the cross-section A (equation (4)) would have to be very small to produce a decay of the duration commonly found. Secondly, when a large number of electrons are in traps (the number of trapped electrons is often of the same order as the number of luminescence centres or impurity atoms in the phosphor) the number of empty luminescence centres is large and recombination would be very rapid, and also the number of free electrons would be much less than the number of empty luminescence centres.

In certain circumstances a bimolecular phosphorescence mechanism may be produced by the time of delay of electrons in traps, the time of movement of electrons in the conduction band being negligible. For instance, consider a phosphor containing traps all of which are shallow and which are filled only to a small extent during excitation. The number of empty luminescence centres is then small compared with the number of empty traps and an electron will be trapped and retrapped many times before reaching a luminescence centre. The recombination of electrons and empty luminescence centres is bimolecular and equation (4) applies. The effect of repeated re-trapping is simply to slow down the movement of the electrons through the phosphor. The rapidity of this type of decay increases very quickly with temperature.

EXPERIMENTAL METHOD

The shape of decay curves will, in many cases, depend on the mode of excitation such as excitation by a single pulse of exciting light or by a periodic succession of pulses. In the present experiments the exciting light passed through a sector disk rotating at constant speed so that periodically the phosphor was excited for a few milliseconds by a pulse of light of constant intensity and was then allowed to decay in the dark for several milliseconds. During the period of excitation the intensity of luminescence increased with

time, and in some cases this rise curve provided more information than the decay curve. After a number of periods of excitation and decay the phosphor reached a cyclic condition and the periodic rise and fall of luminescence was recorded by a photocell multiplier, cathode-ray oscillograph, and camera. Oscillograms of this type are shown in figures 4, etc., the ordinate is light intensity and time is the abscissa; except for figures 12 and 17, where the line of zero light intensity is recorded, the curves show only the change in light intensity and not the absolute light intensity.

The source of exciting light was a 1000 W air-cooled high-pressure mercury lamp, the arc being approximately 4 mm. long and 2 mm. wide. 0.36μ radiation was filtered by means of thick nickel oxide glass and a solution of cupric chloride in water. Unless otherwise stated, 0.36μ radiation was used in our experiments. A single-prism quartz monochromator was used to obtain other wave-lengths. Two thicknesses of Wratten 2A filter prevented exciting light entering the multiplier which had caesium surfaces and ten multiplying stages. The voltage from the multiplier was amplified by a simple two-stage resistance capacity coupled triode amplifier using Osram H 30 valves giving an amplification of 60 in each stage. The time constants were adjusted so that the amplifier responded to a wide frequency range.

We were able to vary the temperature of the phosphor by means of an apparatus which was essentially a small german silver dewar with a large demountable quartz window on the outside. This window could be removed and the phosphor could then be placed on a copper sheet (rhodium plated to avoid contamination of the phosphor) soldered on the outside of the inner tube of the dewar. A single layer of phosphor grains was attached to the rhodium surface by a trace of glycerol. Evacuation of the dewar prevented the phosphor being chemically attacked by the atmosphere or becoming coated with ice when at low temperatures. The inner tube of the dewar could be filled with liquid oxygen to cool the phosphor and contained a small electric heater for raising the temperature. The temperature of the copper sheet was measured by a thermocouple and tests showed that the phosphor, *in vacuo*, was at the same temperature as the copper.

In some experiments, to obtain greater accuracy of measurement, a Becquerel type of phosphoroscope was used. A disk with a radial slot near the outer edge and a quadrant sector of larger radius were mounted in contact on the same axis and rotated together at a constant speed. The exciting light passed outside the disk and was cut off by the quadrant sector during a quarter of each revolution of the system. The luminescent light from the phosphor was focused so that, as the system rotated, the light, for an instant, passed through the radial slot into a 12-stage rubidium multi-

plier. By rotating the quadrant sector relative to the disk one could vary the interval between the cutting off of the exciting light and the instant when phosphorescent light entered the multiplier. A mirror galvanometer gave a steady deflexion which measured the mean photocurrent produced by the pulses of light entering the multiplier.

To measure glow curves and the slow growth of fluorescence with time the 12-stage multiplier and galvanometer were used as in previous work (Randall & Wilkins 1945*b, c*).

EXPERIMENTAL RESULTS

(1) *Exponential decays*

The properties of this relatively simple type of decay have already been fairly adequately surveyed (e.g. Randall & Wilkins 1945*a*; Johnson & Davis 1939); we wish to show here how this type of decay can be distinguished from

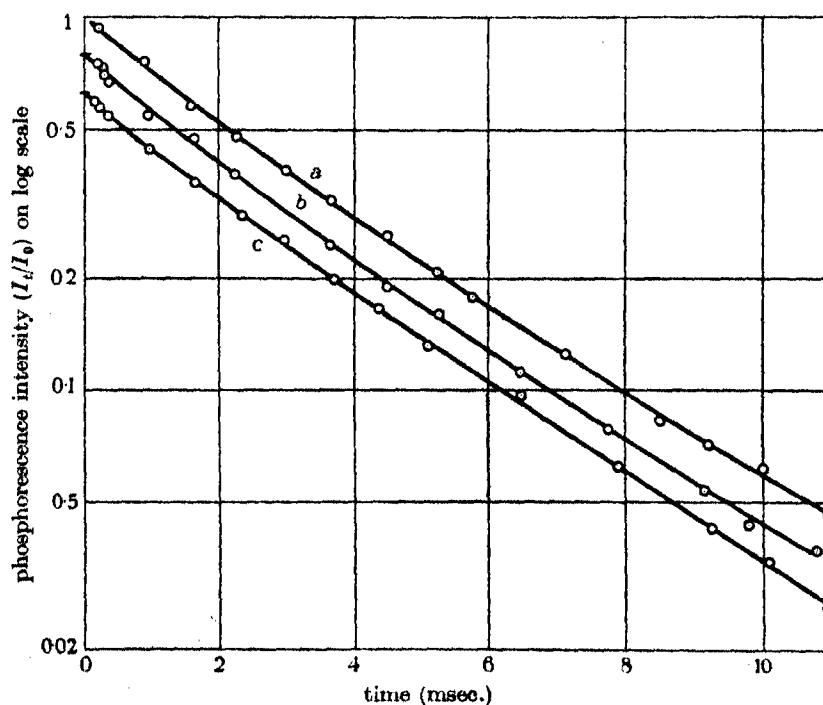


FIGURE 3. Decay curves of ruby measured with Becquerel phosphoroscope. The curves are independent of the wave-length and intensity of the exciting light. Excitation: *a*, 0.36μ radiation; *b*, 0.43μ radiation; *c*, 0.43μ radiation, $\frac{1}{2}$ intensity.

a trap decay and how complex decays arise from the combined action of exponential and trap mechanisms.

Ruby. Although the properties of exponential decays have been explained by the assumption that phosphorescence is the result of the luminescence transition being of a forbidden type, until now experimental proof of this appears to have been lacking. Such proof is now available in the case of

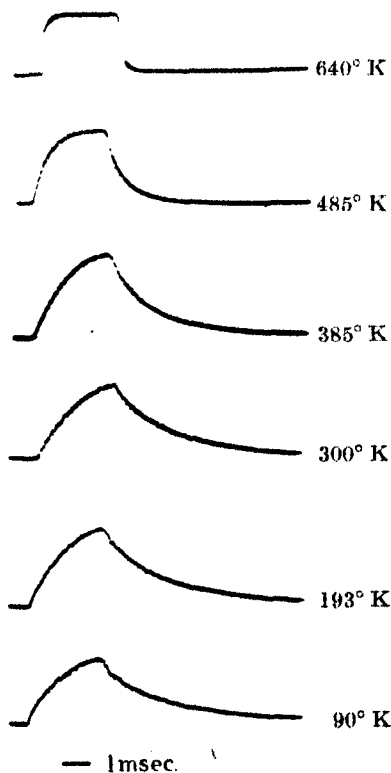


FIGURE 4. Decay of ruby.

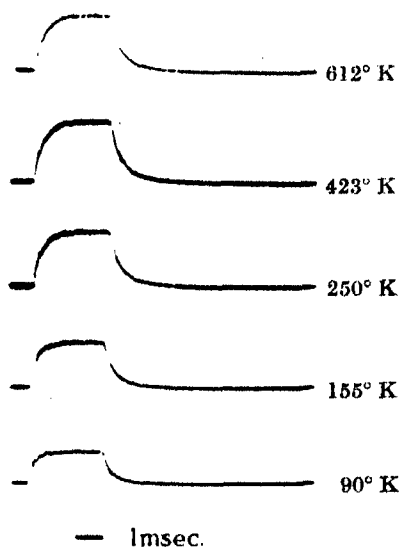


FIGURE 5. Decay of $\text{CaWO}_4\text{-Sm}$.

ruby ($\text{Al}_2\text{O}_3\text{-Cr}$). Deutschbein (1932) has shown spectroscopically that the luminescence is due to a forbidden transition in the Cr^{+++} impurity ion, and we find that this phosphor gives a typical exponential decay (figures 3, 4) of half period ~ 2 msec.

Calcium tungstate-samarium. The decay of this phosphor, like that of the ruby, has the properties described in the theoretical section. The independence of the decay with respect to temperature is well illustrated by this phosphor; the decay scarcely altered from 90 to 600° K, the temperature range covered by the observations (figure 5).

Manganese chloride. Manganese impurity phosphors, to the list of which (Randall & Wilkins 1945 *a*) we can now add CdI_2 (half-period ~ 0.5 msec.) and, when suitably excited, ZnS-Mn with half-period ~ 2 msec., all give approximately exponential decays of half-period 0.5–20 msec. The luminescence and exponential decay is apparently produced by a forbidden transition in the Mn^{++} ion. Randall (1939) has shown that the red lumin-

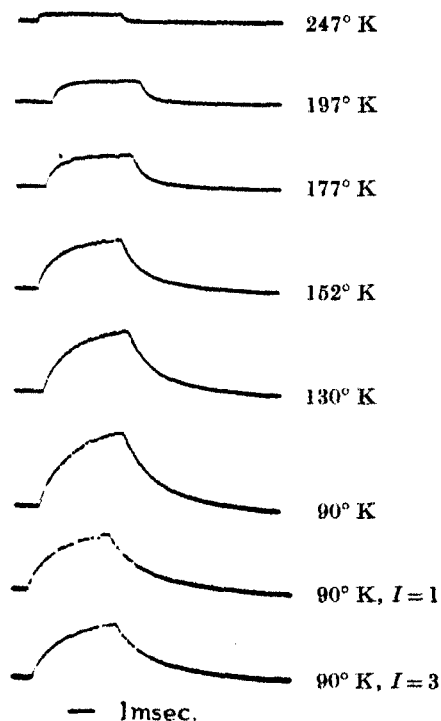


FIGURE 6. Decay of MnCl_2 .

escence, typical of manganese impurity, is also exhibited by pure anhydrous manganese halides at low temperatures. It is interesting to note (figure 6) that the luminescence of these pure salts further resembles that of the manganese impurity phosphors in that the decay is exponential and of half-period ~ 0.5 msec. The decay is independent of intensity of excitation but increases in rapidity as the temperature is raised.

The majority of exponential decays are almost independent of temperature, but the example of MnCl_2 shows that this rule does not always apply. There is, of course, no theoretical reason why the levels in the luminescence centre should not alter with temperature and cause a change in the decay

rate. As the exponential decay is a property of the luminescence transition we may expect, if the decay rate changes, to observe also changes in other properties of the luminescence. We find that, when the decay rate changes, it increases with rise of temperature and, in the same temperature range, the brightness of luminescence decreases; examples are MnCl_2 from 90 to 247°K and ruby at temperatures above 400°K .

(2) *Complex decays produced by both exponential and trap mechanisms. Zinc sulphide manganese*

If a phosphor is photoconducting, excited electrons escape from the luminescence centres and may spend time in traps as well as in the excited state of the luminescence centre; a complex decay curve results which, in the simple case, e.g. $\text{Zn}_2\text{SiO}_4\text{-Mn}$, consists of an initial exponential portion and a tail due to traps (e.g. Randall & Wilkins 1945*b*). In some other cases the exponential and trap decays overlap and are not clearly separable; zinc sulphide manganese is a good example of a phosphor giving this type of decay. A satisfactory analysis of the decay of this phosphor or of a similar decay has not, so far, been available (see discussion of ZnS-Mn , Randall & Wilkins 1945*a*, de Groot 1939*b*).

Kröger (1939*a*) has found that manganese in ZnS-Mn produces absorption bands on the long wave-length side of the ZnS absorption. If the phosphor is excited by light absorbed in these manganese bands there is afterwards no long after-glow or thermal storage of light energy in the phosphor. Absorption of light in the region of ZnS lattice absorption produces long after-glow and thermal storage of light. It was suggested by Kröger that electrons do not leave the manganese ions when light is absorbed by these ions themselves, but, when light is absorbed by the lattice, electrons and positive holes are produced and travel through the crystal to the manganese ions. The travelling electrons may be trapped and produce a long after-glow.

These ideas have been confirmed by our measurements with the Becquerel phosphoroscope of the phosphorescence decay curve of $\text{ZnS} + 2\% \text{Mn}$ excited by different wave-lengths. When excited in a manganese absorption band by 0.43μ radiation the decay is initially approximately exponential and is independent of exciting intensity (figure 7); there is a small tail due to traps. When excited by 0.36μ radiation the decay is non-exponential and intensity dependent, but less intensity dependent than a decay due entirely to traps. Excited electrons spend time both in traps and in the excited state of the manganese ion. The beginning of the decay proceeds at a rate determined by the manganese ion. On shorter wave-lengths the decay is similar,

but there is relatively less trap tail owing to the higher absorption of the exciting light.

To interpret completely the decay curve produced by 0.36μ radiation it is necessary to know the proportion of electrons producing luminescence

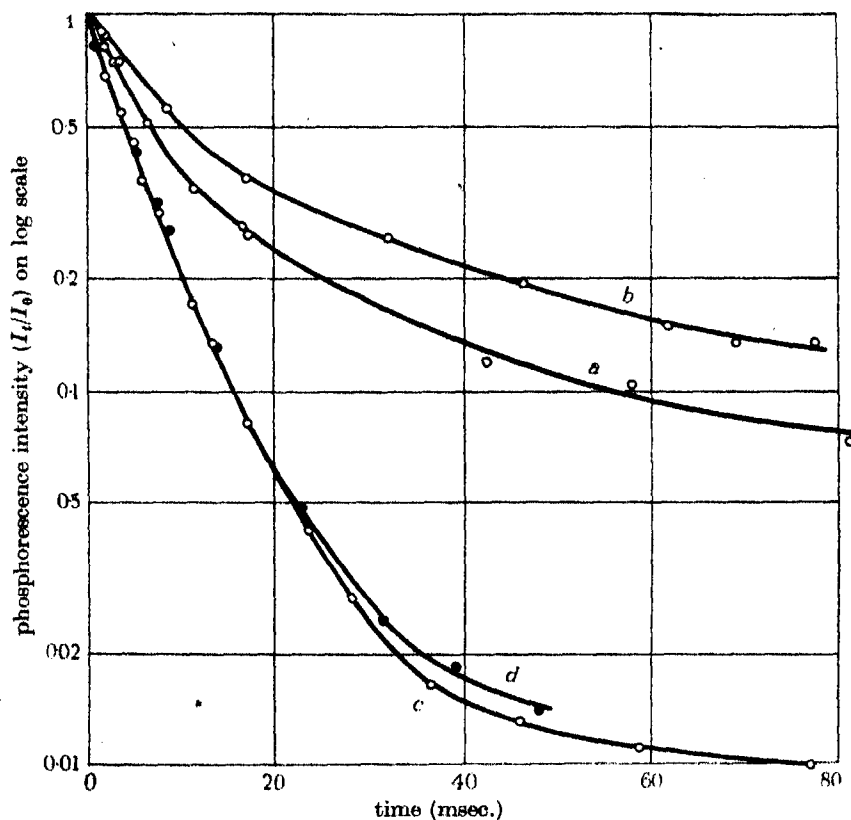


FIGURE 7. Decay curves for ZnS-Mn, measured with Becquerel phosphoroscope. *a*, excited by 0.36μ , $I=9$; *b*, excited by 0.36μ , $I=1$; *c*, excited by 0.43μ , $I=9$; *d*, excited by 0.43μ , $I=1$.

which pass into the conduction band and are available for trapping. Photo-conduction measurements would give us this information, but there is an easier experimental approach which is as follows. The phosphor is warmed to empty all traps and is then cooled in the dark and exposed to a low intensity of exciting light. If all the excited electrons pass into the conduction band the fluorescence rises slowly from zero (e.g. ZnS phosphors other than ZnS-Mn), taking several seconds or minutes to attain equilibrium.

The intensity of fluorescence depends on the relative numbers of empty traps and luminescence centres, and, as the excited electrons fill the traps, the luminescence slowly grows in intensity (Randall & Wilkins 1945*c*). If a fraction f of the excited electrons do not leave the luminescence centres, the luminescence will rise almost instantaneously to f of its equilibrium value.

Figure 8 shows records of the slow growth of fluorescence of ZnS-Mn at 90° K. For 0.43 μ radiation almost all the exciting light is absorbed by the

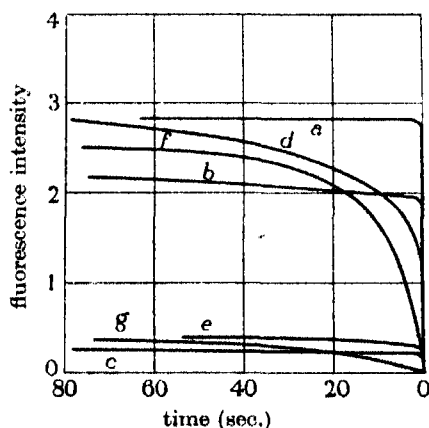


FIGURE 8. Curves showing growth of fluorescence of ZnS-Mn when excited by different wave-lengths. *a*, 0.43 μ ; *b*, 0.36 μ ; *c*, 0.36 μ , low intensity; *d*, 0.33 μ ; *e*, 0.33 μ , low intensity; *f*, 0.31 μ ; *g*, 0.31 μ , low intensity.

manganese ions and the fluorescence rises instantaneously. A few electrons do reach the conduction band and produce a slow filling of traps and a faint phosphorescence tail (figure 7) and some photoconduction (Randall & Wilkins 1945*a*). 0.36 μ radiation is partly absorbed by the manganese ions, thus producing a quick rise of fluorescence, and partly by the ZnS lattice causing a slow rise. The same is true of 0.33 μ radiation. For 0.25 and 0.31 μ all the light is absorbed by the lattice; the absorption coefficient is high, and only the traps on the surface of the crystals are filled first, and the fluorescence rises more rapidly than is the case for 0.36 μ . If we assume the efficiency of production of luminescence is the same when 0.36 μ radiation is absorbed by the lattice or by manganese ions, then at room temperature about 90 % of the excited electrons leave the luminescence centres.

It has been suggested (Mott 1938) that absorption of light in a phosphor may raise an electron to an excited state and thermal energy may then raise some of the excited electrons into the conduction band. We find,

however, that the height of the quick rise on the growth curve of fluorescence for 0.36μ and the rate of filling of traps by 0.43μ radiation are both independent of temperature. Similar results were obtained for cadmium chlorophosphate-manganese. We conclude that thermal energy is not required to raise excited electrons into the conduction band. It would appear that electrons are removed from the luminescence centres as Kröger suggests by light absorption in the matrix and by movement of the positive holes so produced to the luminescence centres where they become trapped.

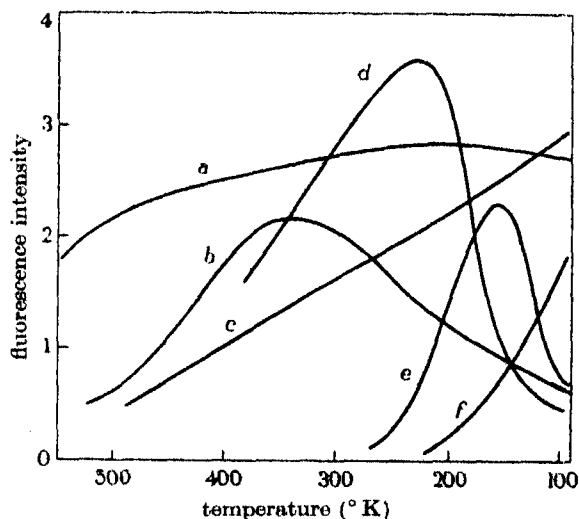


FIGURE 9. Variation of fluorescence brightness with temperature for ZnS-Mn and CdI₂-Mn, excited by various wave-lengths. *a*, ZnS-Mn, 0.40μ , 0.43μ ; *b*, ZnS-Mn, 0.36μ ; *c*, ZnS-Mn, 0.33μ , 0.31μ , 0.25μ ; *d*, CdI₂-Mn, 0.36μ ; *e*, CdI₂ pure, 0.36μ ; *f*, CdI₂-Mn, CdI₂ pure, 0.25μ .

The existence of two types of mechanism for the total process of absorption and luminescence is shown by the different temperature dependence of the fluorescence produced by different wave-lengths of exciting light (figure 9). The fluorescence produced by 0.36μ radiation drops in intensity at low temperatures (Randall 1937). This drop is not due to a lowering of the luminescence efficiency of the transition in the manganese ion. This is shown by the fact that the quick rise on the growth curve for 0.36μ and the fluorescence produced by 0.43μ do not fall at low temperatures. The fluorescence which drops is that part produced by absorption in the lattice. It may be significant that the blue luminescence of some specimens of pure ZnS drops in intensity at low temperatures. A similar effect is found in CdI₂-Mn where

the red fluorescence caused by 0.36μ drops to a small value at low temperatures (figure 9). This effect occurs for 0.36μ only: on shorter wave-lengths the yellow luminescence characteristic of pure CdI_2 is excited in $\text{CdI}_2\text{-Mn}$. The drop in fluorescence is not likely to be due to the action of traps in the phosphor (Gurney & Mott 1939) as there are probably more than a hundred luminescence centres per trap in these phosphors.

(3) *Phosphorescence due to traps*

The following experimental results, mainly for ZnS phosphors, illustrate those characteristic features of a trap decay which have been discussed in the theoretical section. Apart from the effect of intensity and wave-length of exciting light on the decay the experimental results are, in the main, new. The bimolecular theory is shown, on the whole, to be untrue.

Variation of decay rate with number of traps

If we measure the decay curve of phosphors of the same type but containing different numbers of traps we expect I_t/I_0 to increase with the number of traps. Table 1 gives experimental results showing the expected

TABLE 1. RAPIDITY OF DECAY FOR PHOSPHORS CONTAINING
DIFFERENT NUMBERS OF TRAPS

description of phosphor. The composition is given by the ratio of the numbers of metallic atoms present	relative number of traps per unit volume	I_t/I_0 for $t = 3$ msec.
ZnS (blende), 980°C , 10^{-4}Cu	2.4	0.2
ZnS (blende), 970°C , 10^{-4}Cu	1.0	0.3
ZnS (wurtzite), 1100°C , 10^{-4}Cu	1.0	0.3
ZnS.CdS (7:3) (wurtzite), 800°C , 10^{-4}Cu	0.1	0.2
ZnS (blende), 760°C , 10^{-4}Cu	0.1	0.08
ZnS (blende), 980°C (roasted in air 700°C), 10^{-4}Cu	0.05	0.05
ZnS (blende), 600°C , 10^{-4}Cu	0.05	0.02
ZnS, CdS (1:1), 1100°C , 10^{-4}Cu	0.002	0.02

correlation between I_t/I_0 and the number of traps. The correlation is only rough, this probably being due to the fact that the various phosphor specimens contain different numbers of luminescence centres and have different absorption coefficients for the exciting light. The decay will depend on these factors as well as on the number of traps. To find the relative number of traps per unit volume of the phosphor, the height of the glow curve was measured at the appropriate temperature and divided by the fluorescence brightness of the phosphor. In this way allowance was made for the fact that the phosphors had different colours of emission and different efficiencies for converting absorbed energy into luminescence.

Variation of decay rate with temperature

If the glow curve of a phosphor consists of a single peak or several well-separated peaks we expect the decay curve to vary markedly with temperature, I_t/I_0 increasing and decreasing as the glow curve rises and falls.

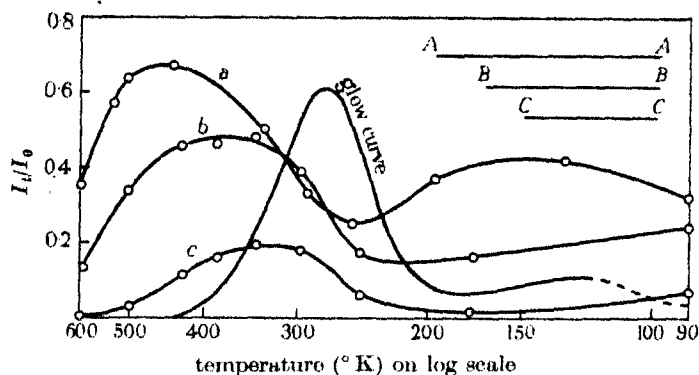


FIGURE 10. ZnS, CdS-Cu. Correlation of glow curve with variation of decay rate. The theoretical displacement of the I_t/I_0 curves a ($t=0.3$ msec.), b ($t=2$ msec.), c ($t=14$ msec.) in relation to the glow curve is given by AA , BB , CC ($s=10^8$ sec. $^{-1}$).

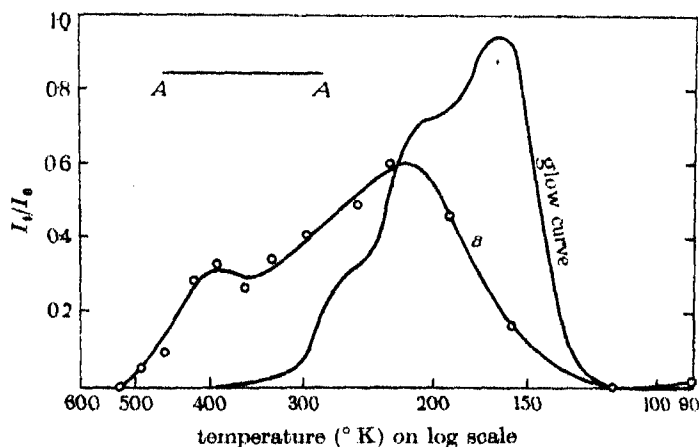


FIGURE 11. Pure ZnS. Correlation of glow curve with variation of decay rate. Curve a is for I_t/I_0 when $t=3$ msec. and AA shows the theoretical shift of curve a with respect to the glow curve.

This effect is well illustrated by figure 10, which shows results for a phosphor ZnS. CdS (87:13) 10^{-4} Cu, Levy West powder E. The decay curves are shown in figure 12. The curve of I_t/I_0 is, as expected, displaced relative to

the glow curve to an extent depending on the value of t at which I_t is measured. The variation of I_t/I_0 is least for small values of t ; this may mean that the initial stages of the decay are not determined by traps.

Similar changes in the decay curve with temperature are shown by many other ZnS phosphor specimens; as an example figures 11 and 13 show results for a pure ZnS phosphor luminescing blue (wurtzite made at 1500° C). The change in the decay curve as the temperature is raised from 90 to 156° K is specially marked and is due to the rapid change in height of the glow curve in this temperature range. In many sulphide phosphors the glow curve is of roughly constant height over most of the temperature range in which luminescence can be excited and hence the decay curve does not vary markedly with temperature, e.g. observations on decay of ZnS-Cu (probably wurtzite) from 295 to 703° K by de Groot (1939*a*).

Change in shape of decay curve with temperature

The shapes of the decay curves in figure 12 are similar to those of the theoretical curves in figure 2. The decay curve at 175° K in figure 12 corresponds with the theoretical decay curve at temperature T_a (figure 2), the

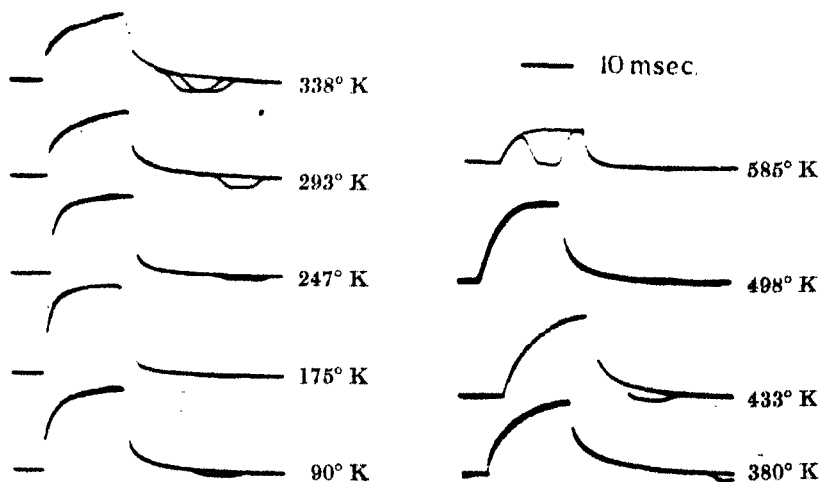


FIGURE 12. Decay curves of ZnS.CdS-Cu. The position corresponding to zero light intensity on the oscillograms is given by an extra rotating sector which rotates in front of the photocell and produces a dent in the decay curve.

rise and decay being rapid. At 247° K, corresponding to T_b , the slow component of decay is appreciable. At 293° K, corresponding to T_c , the slow component of decay is pronounced, but a rapid component is also present and the change of luminescence intensity during the cycle of excitation

and decay is reduced (see also figure 13 curve for 156° K). When 380° K is reached the rapid component of rise and decay is very small and has disappeared at 433° K. At higher temperatures the decay accelerates and at 498° K, corresponding to T_d , the phosphorescence decays to zero in the interval between the periods of excitation. At higher temperatures, 585° K and T_e , the decay becomes more rapid and the brightness of luminescence decreases.

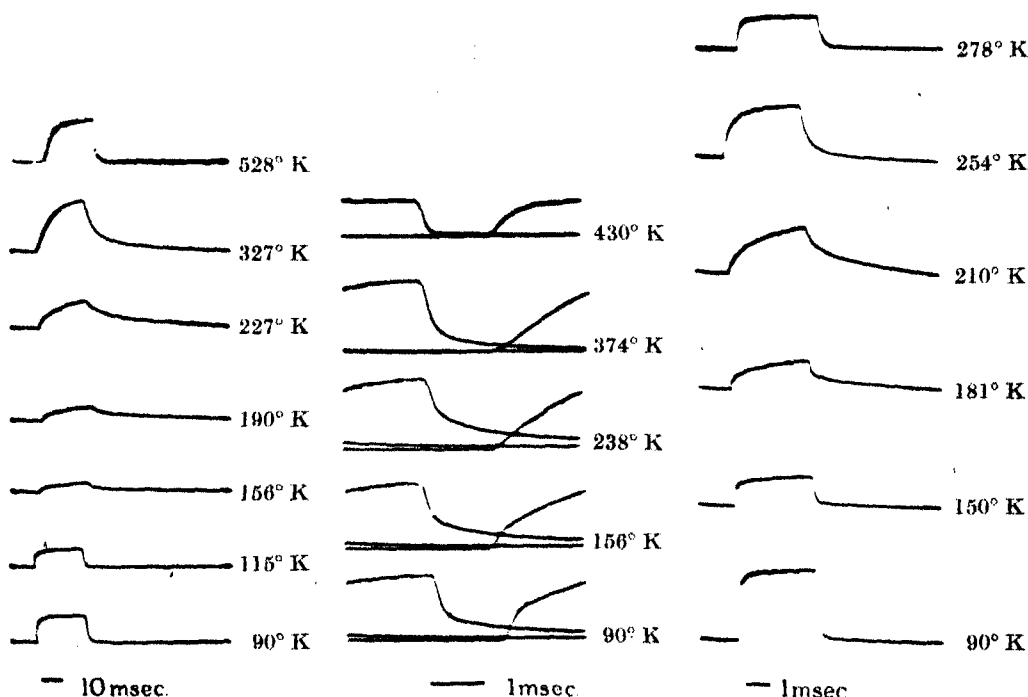


FIGURE 13. Decay of pure ZnS. Gain increased for curve at 528° K.

FIGURE 14. Decay of ZnS-Ag. Gain altered to keep curves approximately constant in height.

FIGURE 15. Decay of ZnO.

The curves of figure 16 for $\text{ZnS} + 10^{-4} \text{ Cu}$ (blende) show well the development of an inflexion on the curve of the rise of fluorescence with time at high temperatures (theoretical curve at T_d , figure 2). At the temperature 90° K, below the maximum of the glow peak, there is a marked rapid component on the rise and decay curves; this component disappears at 327° K, above the maximum of the glow peak, and at 482° K there is a well-developed inflexion on the rise curve. The fourth decay curve (figure 16) is for another phosphor and shows the same kind of inflexion, a small amount of exciting

light being allowed to pass into the multiplier so as to mark the point at which the rise curve begins. Many other ZnS phosphors, e.g. ZnS-Ag and ZnS. CdS-Cu, showed similar inflexions at high temperatures.

The decay curves in figure 14 for ZnS-Ag + 10^{-6} Ni (Levy West phosphor) indicate that there is a physical distinction between the rapid and the slow components of the decay curve at low temperatures. At 90° K the rise and decay consists of two parts, a rapid and a slow component separated by a discontinuity. The time base of the recording oscillograph made three sweeps in the period between successive excitations of the phosphor. The rapid component disappears in a time less than 0.15 msec., the time taken for the exciting light to be cut off by the rotating sector. At 156° K the rapid component is reduced and has disappeared at 238° K. At 374° K the decay is more rapid and there are signs of an inflexion on the rise curve. At higher temperatures the decay becomes still more rapid. The glow curve of this phosphor is similar to that in figure 11 except that the curve has an appreciable height at 90° K.

We attribute, in the theoretical section, the rapid component of decay at low temperatures, such as 90° K, to the fact that there are at these temperatures few empty traps to delay an excited electron and many empty luminescence centres which may capture electrons directly. A rapid decay of the type observed might also be due to excited electrons not leaving luminescence centres at low temperatures. But curves of the slow growth of fluorescence at 90° K show little quick rise of fluorescence and indicate (see experimental section on ZnS-Mn) that the majority of excited electrons do leave the luminescence centres and are thus available for trapping. For the impurity-activated ZnS phosphors the quick rise is (if present) less than 1 % of the total fluorescence, and for the pure ZnS phosphor and the ZnO phosphor the quick rise is certainly less than 10 % of the total.

Decays due to traps are not characteristic of zinc sulphide phosphors only; what observations we have made indicate that similar phenomena take place in zinc oxide, silver chloride, thallous chloride, and alkaline earth sulphides. Figure 15 shows decay curves for a ZnO phosphor (made at 1000° C). The glow curve consists of a double peak extending from 90 to 200° K with maxima at 190 and 130° K. The decay curve changes with temperature in the expected manner.

The curves of figure 17 show the increase in rapidity of the decay with intensity of excitation and are typical of ZnS phosphors of all kinds. Also, the rapidity of decay increases as the wave-length is decreased because short wave-lengths are more heavily absorbed by the phosphor than longer wave-lengths (de Groot 1939a).

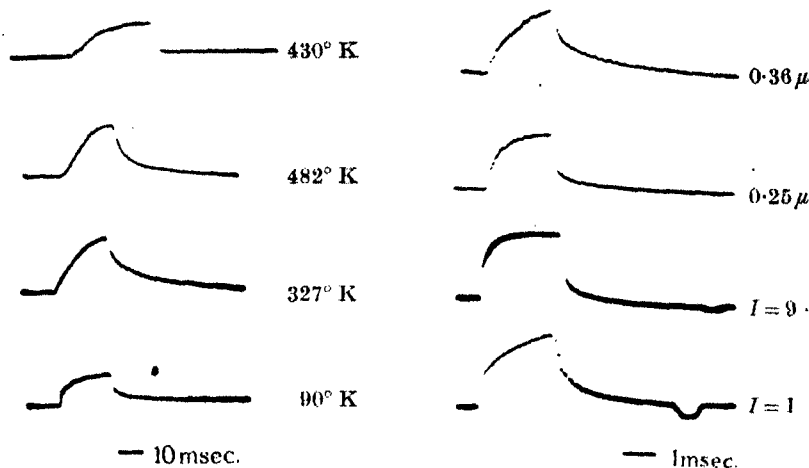


FIGURE 16. Change in shape of ZnS-Cu (blende) with temperature. The top curve is for a different specimen.

FIGURE 17. Change in shape of decay curve of ZnS-Cu with intensity and wave-length of excitation, at 293° K.

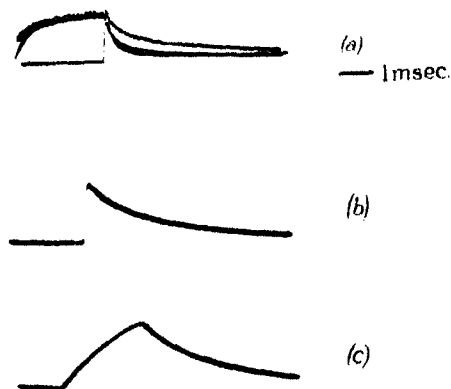


FIGURE 18. *a*, decay of pure ZnS phosphor for different times of excitation at 293° K. Top curve, prolonged excitation, $I = 1$; middle curve, prolonged excitation, $I = 9$; bottom curve, short time of excitation, $I = 9$; so adjusted that the initial phosphorescent intensity is the same as for the curve $I = 1$. The exponential decay of the ruby is unaffected by altering the time of excitation, *b*, ruby decay, short excitation; *c*, ruby decay, prolonged excitation.

Variation of decay curve with time of excitation

According to trap theory the initial decay should become more rapid when the time of excitation is reduced. In figure 18 the curve for short excitation does, in fact, fall slightly below the middle curve which is for prolonged excitation. The bimolecular theory, however, predicts that, for short excitation, the curve should be less steep and should coincide with the curve for less intense excitation, i.e. the top curve in figure 18. Hence experiment contradicts bimolecular theory and supports trap theory.

CONCLUSION

The chief result of this work has been to show that non-exponential decays of variable rate, such as the decay of ZnS-Cu, are due largely to the time electrons spend in traps in the phosphor. The main experimental evidence for this is as follows:

(1) The rapidity of decay of different phosphor specimens of the same type depends mainly on the number of traps in the phosphor.

(2) The way the rapidity of decay and the shape of the decay curve vary with temperature is determined by the glow curve.

The main experimental results indicating that the bimolecular theory should be rejected in favour of trap theory are as follows:

(1) The decay curve is not a function only of the initial phosphorescence intensity as predicted by the bimolecular theory.

(2) The rapidity of decay does not increase steadily with temperature as predicted by the bimolecular theory but may increase and decrease in an irregular manner with temperature.

It is worth noting that we have not needed to make any assumptions about the nature of the traps and luminescence centres in order to account for many of the properties of the decay curve. But to account for the presence of rapid and slow components on the decay curve at low temperatures and of an inflexion on the rise curve at high temperatures, we need to assume that excited electrons are completely removed from luminescence centres. It is clear, however, that any picture of phosphorescence mechanism must necessarily be tentative until the chemical and physical nature of phosphors and the traps and luminescence centres they contain is understood. Further advances in studying the physical properties of phosphorescence will depend, to a large extent, on the chemical production, by controlled means, of phosphors with different concentrations of electron traps and impurity centres.

We wish to thank Professor M. L. Oliphant, F.R.S., who provided encouragement and facilities for this work. We are specially indebted to Dr J. T. Randall for his constant interest and helpful advice and to Professor R. Peierls and Dr K. Fuchs who have helped by discussing the theoretical aspects. Many of the phosphors were made by Dr A. F. Wells and by Dr Randall. Our thanks are due to the Warren Research Fund Committee of the Royal Society for a grant to one of us (G. F. J. G.) and to the Department of Scientific and Industrial Research for a grant (to M. H. F. W.). We also wish to thank Marconi Wireless Telegraph Company for the caesium multiplier used in our experiments.

REFERENCES

- Becquerel, E. 1867 *La Lumière, ses Causes et ses Effets*. Paris.
 Deutschbein, O. 1932 *Ann. Phys., Lpz.*, **14**, 712.
 Groot, W. de 1939*a* *Physica*, **6**, 275.
 Groot, W. de 1939*b* *Physica*, **6**, 393.
 Gurney, R. W. & Mott, N. F. Luminescence, A General Discussion 1939 *Trans. Faraday Soc.* **35**, 72.
 Johnson, R. P. 1939 *J. Opt. Soc. Amer.* **29**, 387.
 Johnson, R. P. & Davis, W. L. 1939 *J. Opt. Soc. Amer.* **29**, 283.
 Kröger, F. A. 1939*a* *Physica*, **6**, 369.
 Lenard, P., Schmidt, F. & Tomaschek, R. 1932 *Handb. Exp. Phys.* **23**.
 Lewschin, W. L. & Antonow-Romanowsky, W. W. 1934 *Phys. Z. Sowjet.* **5**, 379.
 Mott, N. F. 1938 *Proc. Roy. Soc. A*, **167**, 384.
 Pohl, R. W. 1937 *Proc. Phys. Soc.* **49** (extra part), 3.
 Randall, J. T. 1937 *Proc. Phys. Soc.* **49** (extra part), 52.
 Randall, J. T. 1939 *Proc. Roy. Soc. A*, **170**, 272.
 Randall, J. T. & Wilkins, M. H. F. 1945*a* *Proc. Roy. Soc. A*, **184**, 347.
 Randall, J. T. & Wilkins, M. H. F. 1945*b* *Proc. Roy. Soc. A*, **184**, 365.
 Randall, J. T. & Wilkins, M. H. F. 1945*c* *Proc. Roy. Soc. A*, **184**, 390.
 Riehl, N. 1939 *Trans. Faraday Soc.* **35**, 135.
 Seitz, F. 1938 *J. Chem. Phys.* **6**, 150.

On the catalytic cyclization of aliphatic hydrocarbons. I

BY E. F. G. HERINGTON AND E. K. RIDEAL, F.R.S.

Department of Colloid Science, University of Cambridge

(Received 3 September 1941)

The aromatization of a paraffin proceeds through the formation of a mono-olefine which is held on the catalyst by two-point contact. The promoter action of the central atoms in heteropoly acids of molybdenum can be attributed to the stabilization of the molybdenum dioxide produced during the hydrogen pretreatment. The alumina used as a support for molybdenum catalysts plays the same role. In the reduction of phenol to benzene by the heteropoly acids of molybdenum the metal is the catalyst.

The variation of yield of the total aromatic hydrocarbon with different paraffin feeds can be calculated by assuming two point contact.

A yellow compound (fulvene or a polymer of fulvene) is for given conditions produced in a constant ratio to the aromatic hydrocarbon.

The loss of activity of the catalyst with use is a result of the polymerization to giant molecules of the hydrocarbons adsorbed on adjacent catalyst centres. This decay rate may be diminished by decreasing the concentration of active centres on the surface.

It is well known that on the passage of paraffins with chains of more than six carbon atoms, or the corresponding mono-olefines over suitable catalysts, aromatic hydrocarbons are formed. The thermodynamics of these reactions have been reviewed by Taylor & Turkevich (1939). It has also been shown that certain oxides are the most effective catalysts, and from the work of Pitkethly & Steiner (1939) it was concluded that a mono-olefine was the necessary intermediary in the reaction where paraffins were concerned. The mechanism of the formation of the olefine and of any necessary migration of the double bond along the chain prior to ring closure has been shown to proceed through the half-hydrogenated state by Twigg & Rideal (1939, 1940, 1941).

Whilst we now have some idea as to the mode of attachment of the hydrocarbon to metals, little is known about the oxide catalysts which operate so successfully in this process of ring closure. Again the fact that the yield and nature of the cyclic hydrocarbons produced varies with the nature of the initial aliphatic material, as revealed by the exhaustive analyses of Hoog, Verheus & Zuiderweg (1939), merits some inquiry. Finally, it is observed that this reaction is always accompanied by the phenomenon of

catalyst decay. It has been shown that this is due to poisoning and that the poison which can be burnt off to regenerate the catalyst contains both carbon and hydrogen (Obland, Marschner & Heard 1940). It appears to be formed during the reaction, and it has been found possible to examine the kinetics of the poisoning process and make certain inferences as to the mechanics of the reaction.

THE ACTIVITY OF CATALYSTS FOR CYCLIZATION

The catalysts were employed in the form of granules *ca.* 5 mm. in diameter and were reduced in hydrogen for 6 hr. at 475° C before use. The operating temperature was maintained at 475° C and a 50:50 mixture of *n*-heptane and *n*-heptene was used as starting material. The liquid was passed into the reaction vessel by means of an electrolytic doser, the gases liberated by electrolysis forcing the liquid at a steady rate through a fine capillary. Provision was made for the employment of hydrogen or nitrogen as carrier gas. The contact time during the experiment was 12 sec. The toluene content of the treated material was determined from measurements of the refractive index corrected where necessary for change in olefine content which was measured by the McIlhiney Bromine Number method. In table 1 are given the initial activities (T_0) for a series of catalysts as disclosed by the percentage conversion of the reactants to toluene.

TABLE 1. ACTIVITIES OF CATALYSTS

catalyst	T_0
Al_2O_3	0
MoO_3	0
10 % MoO_3 on Al_2O_3	24
aluminium phosphate	0
20 % MoO_3 on aluminium phosphate	4
Cr_2O_3	41
10 % MoO_3 on Cr_2O_3	28
ammonium alumino-molybdate (6 acid)	26
ammonium thorio-molybdate (12 acid)	38
phospho-molybdic acid	17
ammonium chromo-molybdate (6 acid)	29
ammonium ferro-molybdate (6 acid)	16
ammonium cobalto-molybdate (6 acid)	27
ammonium nickelo-molybdate (6 acid)	17
ammonium phospho-vanado-molybdate	41
ammonium vanado-molybdate	32

It will be observed that in addition to chromium oxide a number of other catalysts possess high activities. Molybdenum oxide deposited on alumina

is relatively active but the two components are themselves inactive. Alumina must thus serve other purposes than as a mere support. It serves a dual purpose, both as a support for the active aromatizing molybdenum oxide catalyst and to form an alumino-molybdate compound in which the oxide is not readily reduced below the state of MoO_2 to inert metal. This view is supported by the following experiments in which molybdic oxide prepared from ammonium molybdate was reduced in hydrogen and was compared with the reduction of ammonium alumino-molybdate. The rate of loss in weight of the two samples is given in table 2.

TABLE 2. REDUCTION OF CATALYSTS

time of heating hr.	temp. °C.	catalyst	% Mo remaining as MoO_2
30	475	Mo } Al-Mo }	all as MoO_2
15	550	Mo Al-Mo	14 % 94 %
48	575	Mo Al-Mo	2 % 48 %

Whilst ammonium molybdate is readily reduced beyond the stage of MoO_2 , that containing aluminium is but slowly reduced to metallic molybdenum. This reduction to the metal must occur on the surface of the ammonium molybdate catalyst even at 475°C , but is very considerably retarded in the complex.

Ammonium thorio-molybdate (an active catalyst) shows a similar behaviour to the alumino complex, while ammonium cupro-molybdate (catalytically inactive for cyclization) is readily reduced at 475°C to pyrophoric molybdenum metal.

Similar conclusions can be drawn from some recent work by Kazonsky and other workers (1940*a*, 1940*b*) in which vanadium oxides supported on alumina are found to be active in contrast to the inactivity of the unsupported oxides. In this case the active catalyst is probably V_2O_3 , for both V_2O_3 and MoO_2 have two unpaired electrons per metal atom when compared with the fully saturated oxides V_2O_5 and MoO_3 . In this connexion it is significant to note that the reduction in hydrogen of V_2O_5 to a lower oxide has been observed at a temperature of *ca.* 500°C (Cameron 1938).

Since chromium and molybdenum occupy the same group in the periodic table, we may infer by analogy that the active chromium oxide catalyst

should be CrO_3 . Hydrated chromium sesquioxide is readily oxidized to a higher oxide (see for example Cameron, Harbard & King 1939). Turkevich, Fehrer & Taylor (1941) have reported the preparation of chromium oxide catalysts of high activity by the partial reduction of chromic acid. Mixing a solution of chromic acid and chromium hydroxide gives a higher oxide than Cr_2O_3 (Dunncliff & Kotwani 1931; Bhatnagar, Cameron, Harbard, Kapur, King & Prakash 1939).

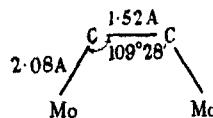


FIGURE 1

There is much evidence (see p. 260) that the olefines are held by two-point contact. The adsorption on MoO_3 may take place as in figure 1.

Taking the values $\text{Mo}-\text{C} = 2.08 \text{ \AA}$ as determined from the hexacarbonyl by Brockway, Ewens & Lister (1938) and the distance $\text{C}-\text{O}$ as identical with that in the ethers, viz. 1.44 \AA according to the measurements of Sutton & Brockway (1935), then assuming the tetrahedral angle $\alpha = 109^\circ 28'$ we obtain the values $\text{Mo}-\text{Mo} = 2.90 \text{ \AA}$, whilst X-ray measurements give

$$\text{Mo}-\text{Mo} = 2.79 \text{ \AA},$$

$$\text{and } = 3.71 \text{ \AA}.$$

Thus two-point adsorption of an ethylene linkage can occur on the Mo ions in MoO_3 and the heteropoly acids, since X-ray examination by Kingman (1937) revealed the fact that the MoO_3 structure was produced when heteropoly acids containing molybdenum were reduced in hydrogen.

In the active catalysts two unpaired electrons are present in the metallic ion and mere geometric fit is not a sufficient condition for catalytic activity. The spacings in molybdenum itself are $\text{Mo}-\text{Mo} = 3.14$ and 2.20 , so that a fit on the long distance might be just possible, but the metal appears to be inert for aromatization. Similarly, the $\text{Al}-\text{Al}$ distance in alumina is correct for adsorption although the material is catalytically inactive.

Some observations of Kingman (1937) on the effect of promoters on molybdenum catalysts used for the hydrogenation of phenol to benzene receives an explanation in the light of the above. Kingman found that unpromoted molybdenum oxide caused reduction of phenol, hence metallic molybdenum must be the catalyst. It follows that the catalysts containing copper together with molybdenum will possess a high activity for phenol reduction, as is indeed found.

Since the mechanism of reduction of phenol does not involve adsorption of an olefine link by two-point contact, but two single links for the phenol and hydrogen respectively, we may conclude that in general an effective

catalyst for aromatization will be found to be a poor catalyst for the reduction of phenol and vice versa. This is in agreement with experiment.

THE PRODUCTS OF CATALYTIC CYCLIZATION

The variation of yield of aromatic hydrocarbons when different initial paraffins are used can be predicted on the assumption that the mono-olefine formed as intermediary is held by two-point contact and that the double bond can migrate rapidly to all positions in the chain, as occurs on metallic catalysts. Thus hexane, after the loss of two hydrogen atoms, can be adsorbed in any one of the configurations shown in figure 2*a, b, c*. With six carbon atoms in the chain there are five pairs to choose from for the depicted possible adsorptions. In these only (*a*) can lead to ring closure, as in the other cases the ends of the chain are away from the catalyst surface. For the adsorption as in figure 2*a* let the probability that carbon (1) will react with carbon (6) to close the ring be p , then the total probability that hexane will react to form an aromatic hydrocarbon is $2p/5 = 0.4p$. In the case of normal paraffins (C_nH_{2n+2}), where n is greater than 9, certain central atoms can react in two ways, e.g. see figure 3, where carbon atom 1 can react with 6 or 10 with 5. The probability that the pair (5, 6) can react is therefore $2p$.

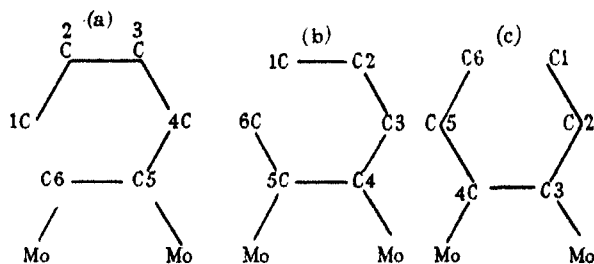


FIGURE 2

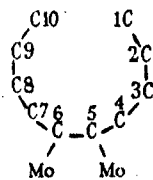


FIGURE 3

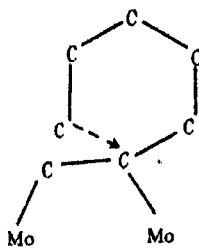


FIGURE 4

If it be assumed that links such as shown in figure 4 do not yield an aromatic hydrocarbon, the probability for the normal paraffin (C_nH_{n+2}) can be shown to be

$$P = \frac{2(n-5)p}{n-1}.$$

For paraffins other than normal the values of P can best be calculated for each separate case, as, for example, with 2,5 dimethyl hexane (see figure 5).

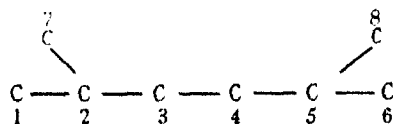


FIGURE 5

There are $(n-1) = 7$ ways of choosing adjacent pairs of carbon atoms. If carbon atom 1 and 2 are attached to the catalyst, then number 1 atom can react in two ways (with 8 or 6), i.e.

$$\text{probability for this pair} = 2p.$$

There are four such pairs (1, 2), (2, 7), (5, 6), and (5, 8) and no other cases lead to aromatization, i.e.

$$P = \frac{8p}{7} = 1.14p.$$

For standard conditions of operation the velocities of conversion for the different hydrocarbons investigated will be proportional to the probabilities for reaction, i.e.

$$k_1 = \text{constant} \times P,$$

where

$$k_1^* = \log_{10} \frac{1}{1-\alpha}.$$

α = fraction converted.

In table 3 are given the calculated values compared with the observed as found by Hoog *et al.* (1939). The constant was taken as equal to 0.282. The agreement is satisfactory. The mean deviation for the cases where $P = 0$ is 3.3 %, i.e. mean deviation for all cases, is not much greater than discrepancies due to experimental error or isomerization of feed (mean deviation of all = 4 %).

The argument should be equally applicable to mono-olefine feed materials provided they are compared among themselves. The theoretical ratio of

the velocity of conversion of heptene to conversion of hexene should be $0.66p/0.40p = 1.7$, while in the present investigation the experimental ratio was found to be 1.8.

TABLE 3. VARIATION OF YIELD OF AROMATICS
WITH DIFFERENT BASE MATERIALS

paraffin	P/p	conversion to aromatic (%)	
		observed	calculated
pentane	0.00	3	0
<i>n</i> -hexane	0.40	20	23
2-methyl pentane	0.00	5	0
<i>n</i> -heptane	0.66	36	35
2-methyl hexane	0.66	31	35
<i>n</i> -octane	0.84	46	42
3-methyl heptane	0.84	35	42
2.5-dimethyl hexane	1.14	52	52
2.2.4-trimethyl pentane	0.00	3	0
<i>n</i> -nonane	1.00	58	48

This treatment does not, however, account for the distribution of the different products, as can be seen from table 4. The experimental values are taken from Hoog *et al.* (1939). The formation of the aromatic hydrocarbon may be accompanied by isomerization. Indeed, the occurrence of isomerization would appear to be proved by an observation of Komarewsky & Coley (1941), who report the formation of 2.6-dimethyl phenol in the cyclization of *n*-octyl alcohol. This product can be formed only by isomerization.

TABLE 4

base material	aromatics which can be formed by direct ring closure	observed	
		observed	calculated
<i>n</i> -hexane	benzene	~100	100
<i>n</i> -heptane	toluene	> 95	100
2-methyl hexane	toluene	~100	100
<i>n</i> -octane	ethyl benzene	—	66
	<i>o</i> -xylene	> 76	33
3-methyl heptane	ethyl benzene	5	33
	<i>o</i> -xylene	35	33
	<i>p</i> -xylene	55	33
2.5-dimethyl hexane	<i>p</i> -xylene	~80	100
<i>n</i> -nonane	<i>n</i> -propyl benzene	—	50
	<i>o</i> -methyl ethyl benzene	~72	50

The isomerization would appear to take place only once; for example, no trimethyl benzene has been reported in the product from nonane. It would probably be better to replace the rule, 'the aromatics formed in the cyclization of aliphatic hydrocarbons contain the shortest possible side chain' (Hoog *et al.* 1939), by the rule, 'the main product contains one methyl group and the remaining carbon atoms in excess of six in another side chain'. The observed distribution of products may result from the preferential cracking of certain of the primary products.

SECONDARY PRODUCTS OF THE REACTION

If in figure 2 above, linkage occurs between carbon atom 2 and carbon 6, a five-membered ring will result which by successive dehydrogenations may yield a derivative of cyclopentadiene or fulvene.

The yellow compound in the product was identified as a fulvene by a study of the material formed by the circulation of toluene over an ammonium chromo-molybdate catalyst at 475° C until the catalyst was poisoned for the formation of coloured products. The excess toluene was removed by distillation. The residue partly solidified, and was separated into three fractions by the use of ether, and methyl alcohol-water mixtures as solvents. The most insoluble fraction in ether was identified as anthracene by its melting-point and oxidation to oxanthranol. The fraction readily soluble in ether, but more insoluble in methyl alcohol-water mixtures, was found to be stilbene. The yellow constituent remained in the most soluble fraction and was contaminated with the other two hydrocarbons. This mixture gave the following reactions which have been described for fulvenes by Thiele (1900):

- (1) The yellow compound was an oil.
- (2) The compound was remarkably volatile in ether vapour.
- (3) If to the solution in acetic acid sulphuric acid was added, a white precipitate separated which became reddish brown on the adding of more sulphuric acid.
- (4) Sulphuric acid alone gave a deep red substance which blackened on the addition of water.

In addition to these reactions dimethyl fulvene was prepared, and it was found that the addition of bromine in carbon tetrachloride gave a colourless bromide which became green on the addition of concentrated sulphuric acid. The material from the circulation of toluene behaved similarly. The benzene ring of toluene must have opened and closed again to form a cyclopentadiene derivative.

The presence of stilbene and anthracene in the product from the circulation of toluene may be readily explained if it be assumed that toluene is adsorbed in the form shown in figure 6. After adsorption the hydrocarbon may undergo dehydrogenation followed by dimerization. If benzene in a stream of hydrogen be passed over the catalyst no yellow product is formed, but toluene gives a yellow product and ethyl benzene produces not only an intensely yellow product, but styrene in addition. Hexane gives a yellow product. This apparent irreversibility in the case of benzene can be explained on the assumption of two-point contact.

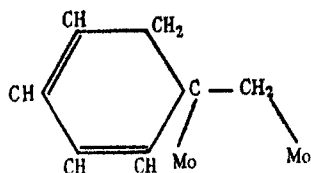


FIGURE 6

Thus if we take the distances

C—C in benzene ring as 1.33 Å,

C—C in side chain as 1.52 Å,

we note that in the ring there is no correct distance for initial adsorption. However, if adsorption takes place involving a side-chain carbon atom as in figure 6, scission of the ring may follow.

The product obtained normally from the aromatization of a paraffin probably contains a coloured polymer of the fulvene. If these views are correct, the concentration of yellow compound in the product should closely follow that of the aromatic.

To measure the concentration of the yellow compound, use was made of the observation that the colour could be matched with that of a K_2CrO_4 solution. For convenience a chromate number (*I*) was defined as the cm. of 0.001 M K_2CrO_4 solution equivalent in colour intensity to 1 cm. of the product.

Passage of heptane through the catalyst tube at 475° C in the absence of a catalyst gave a colourless product. Table 5 shows that, as expected, the ratio *I/T* is constant for the passage of heptane over a reduced chromolybdate catalyst with a contact time of 12 sec.

TABLE 5. TOLUENE CONCENTRATION (*T*) AND CHROMATE NUMBERS (*I*)

time sample taken in hours	toluene (<i>T</i>)	chromate number <i>I</i>	<i>I/T</i>
0.75	13	1.4	0.11
2.62	9	0.9	0.10
4.00	4	0.3	0.08
5.50	2	0.2	0.10

POISONING OF THE CATALYST

The formation of reactive hydrocarbons on adjacent pairs of active centres might be expected to result in linkage of the carbon chains across from one centre to an adjacent one, thus producing a polymer which would be held by four-point contact. Repetitions of such linkage would lead to giant molecules covering the surface. An active centre may therefore either catalyse the formation of an aromatic hydrocarbon, or catalyse the formation of a poison, i.e.

$$\frac{dT}{dt} = -kT,$$

where T = aromatic produced at any instant, t = time; i.e. $T = T_0 e^{-kt}$ giving the variation of aromatic yield with time. Figure 7 shows that the decay curves follow this law.

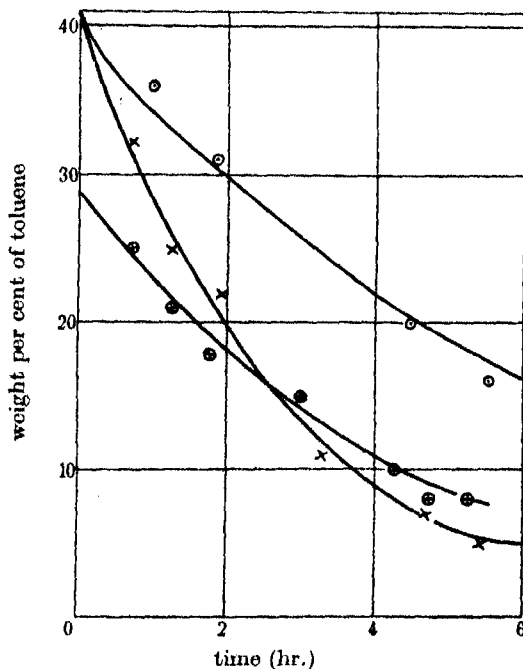


FIGURE 7. Decay of activity with time; 50-50 heptane-heptene feed; temperature 475°C; contact time 12 sec. Points observed, curves calculated. \odot Cr_2O_3 , \times vanado-phospho-molybdate. \odot chromo-molybdate.

Cross-linkages will be rarer the more widely spaced are the active centres. Table 6 substantiates this view. The values in columns 2 and 3 are the values of T_0 and k (log to the base 10, time t in hours) found experimentally,

column 4 gives the values for CO_2 adsorption per gram of catalyst multiplied by the weight of catalyst used. The values of CO_2 adsorption should be proportional to the available areas and were taken from the paper by Kingman (1937).

TABLE 6. VARIATION OF RATES OF DECAY WITH CONCENTRATION OF ACTIVE CENTRES ON CATALYST

catalyst	T_0	k	area A c.c. of CO_2 adsorbed	T_0/A
chromo molybdate	29	0.10	26.0	1.11
thorio-molybdate	38	0.21	10.7	3.55
phospho-molybdate	17	0.07	33.8	0.50
10 % MoO_3 on Al_2O_3	17	0.07	51.5	0.33

The greater the values of T_0/area given in column 5, the greater the concentration of active centres and, as can be seen, the greater is k .

This mechanism for poisoning can be tested further by the use of 'dilute' catalysts where the complications due to the presence of different elements can be avoided.

Chromium and aluminium hydroxides were co-precipitated by the addition of ammonia to a well-stirred solution of the nitrates. The aluminium was the diluent, the chromium the catalyst. The oxides were dried at 100°C , broken into suitable fragments, and reduced in hydrogen for 6 hr. at the reaction temperature of 475°C . The activity was tested by measuring the conversion to toluene when heptane was passed using a contact time of 12 sec.

Since the throughputs of heptane were constant throughout, the relative activities of the catalysts, assuming the reaction is first order, can be calculated from the equation

$$k_1 = \log_{10} (1/1 - \alpha),$$

where α = fraction of heptane converted to toluene. In figure 8 is plotted k_1 (the initial activities) against α_1 , the calculated fraction of Cr on the surface:

$$\alpha_1 = N_{\text{Cr}}^{\text{I}} / (N_{\text{Cr}}^{\text{I}} + N_{\text{Al}}^{\text{I}}),$$

where N_{Cr} equals the number of chromium atoms and N_{Al} the number of aluminium atoms in the bulk.

Figure 9 shows the decay rates (k) plotted against the initial activities (k_1). Clearly k is smaller for lower initial activities.

A confirmation of the mode of poisoning is afforded by the fulvene concentrations. This substance is probably only one of many hydrocarbons which may poison the catalyst but because of its colour it can be used as an

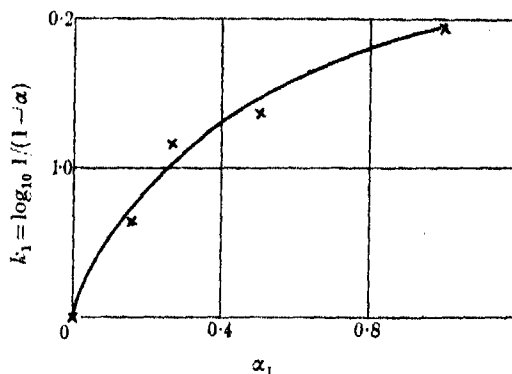


FIGURE 8. Initial activity against fraction of chromium on surface.

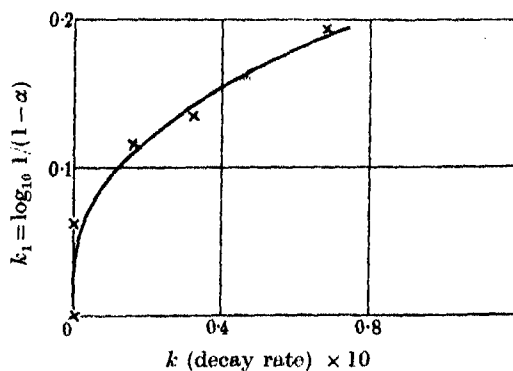


FIGURE 9. Initial activity against decay rate.

indicator. Figure 10 gives the decay rate plotted against I/T (ratio of fulvene to toluene in product). The lower I/T (i.e. the more the fulvene is polymerized) the greater is the decay rate k . The activity k_1 falls off with decreasing values of α_1 much more slowly than would be expected, while at the same time k (decay rate) decreases more rapidly than k_1 . Alumina besides acting as a diluent causes an increase in the available area per gram of catalyst as the values 7.5 and 19.5 c.c. of CO_2 adsorbed per gram of catalyst for $\alpha = 1$ and 0.5 show. This is sufficient to explain the above behaviour.

This work, which forms part of the programme of the Fuel Research Board of the Department of Scientific and Industrial Research, has been

carried out in the Department of Colloid Science, Cambridge. This account is printed by permission of Dr F. S. Sinnatt, Director of Fuel Research. The illustrations are Crown copyright and are reproduced by permission of the Controller of H.M. Stationery Office.

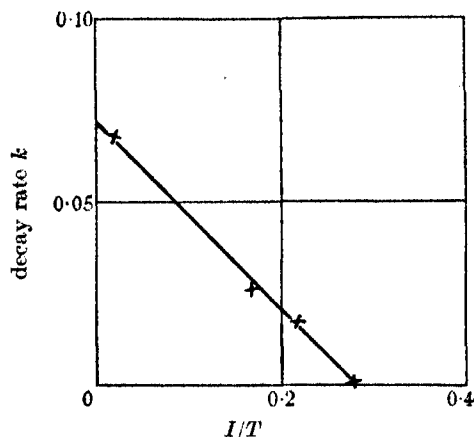


FIGURE 10. Decay rate (k) plotted against I/T for Cr_2O_3 , Al_2O_3 catalysts.

REFERENCES

- Bhatnagar, S. S., Cameron, A., Harbard, E. H., Kapur, P. L., King, A. & Prakash, B. 1939 *J. Chem. Soc.* p. 1433.
 Brockway, L. O., Ewens, R. V. G. & Lister, M. W. 1938 *Trans. Faraday Soc.* **34**, 1350.
 Cameron, A. 1938 Ph.D. Thesis, London.
 Cameron, A., Harbard, E. H. & King, A. 1939 *J. Chem. Soc.* p. 55.
 Dunncliff, H. B. & Kotwani, G. S. 1931 *J. Phys. Chem.* **35**, 3214.
 Hoog, H., Verheus, J. & Zuiderweg, F. J. 1939 *Trans. Faraday Soc.* **35**, 993.
 Kazonsky, Plate, Bulanovo & Zelinsky 1940 *C.R. Acad. Sci. U.R.S.S.* **27**, no. 7, 658-663.
 Kazonsky, Serguienko & Zelinsky 1940 *C.R. Acad. Sci. U.R.S.S.* **27**, 664-669.
 Kingman, F. E. T. 1937 *Trans. Faraday Soc.* **33**, 784.
 Komarewsky, V. I. & Coley, J. R. 1941 *J. Amer. Chem. Soc.* **63**, 700.
 Obland, A. G., Marschner, R. F. & Heard, L. 1940 *J. Amer. Chem. Soc.* **62**, 2066.
 Pitkethly, R. C. & Steiner, H. 1939 *Trans. Faraday Soc.* **35**, 979.
 Sutton, L. E. & Brockway, L. O. 1935 *J. Amer. Chem. Soc.* **57**, 473.
 Taylor, H. S. & Turkevich, J. 1939 *Trans. Faraday Soc.* **35**, 921.
 Thiele, J. 1900 *Ber. dtsch. Chem. Ges.* **33**, 666-673.
 Turkevich, J., Fehrer, H. & Taylor, H. S. 1941 *J. Amer. Chem. Soc.* **63**, 1129.
 Twigg, G. H. 1939 *Trans. Faraday Soc.* **35**, 934.
 Twigg, G. H. 1941 *Proc. Roy. Soc. A*, **178**, 106.
 Twigg, G. H. & Rideal, E. K. 1939 *Proc. Roy. Soc. A*, **171**, 55.
 Twigg, G. H. & Rideal, E. K. 1940 *Trans. Faraday Soc.* **36**, 533.

On the catalytic cyclization of aliphatic hydrocarbons. II

By E. F. G. HERINGTON AND E. K. RIDEAL, F.R.S.

(Received 24 November 1943)

The products formed by the passage of a number of pure hydrocarbons over an aromatizing catalyst have been analysed by a spectroscopic technique. The occurrence of two different types of isomerization has been demonstrated. One may be exemplified by the formation of *p*- and *m*-xylenes from *n*-octane. This isomerization accompanies ring closure and proceeds smoothly at 475° C; it entails the formation and rupture of a subsidiary ring.

The apparent discrepancy between the observations of Hoog, Verheus & Zuiderweg (1939), who report that 2.2.4-trimethyl pentane does not aromatize, and the observations of Green & Nash (1941), who report that a considerable quantity of aromatics is formed, has been shown to arise from the different temperatures employed. The formation of aromatics from substituted pentanes is a general phenomenon; the products from all the isomeric octanes in question have been analysed. The isomerization proceeds through a five-six carbon atom ring mechanism.

In Part I (Herington & Rideal 1945) it was shown that a satisfactory mechanism for the process of catalytic ring closure of aliphatic hydrocarbons on oxide catalysts could be devised on the assumption that the process involved a two-point contact between the catalyst and a pair of carbon atoms in the hydrocarbon, the ring closing by reaction between a carbon atom of the chain which is in the gas phase and one of the adsorbed carbon atoms. By means of this model the dependence of total yield of aromatic hydrocarbon on the structure of the paraffin could be calculated from simple statistical considerations. This treatment, however, did not correctly predict the relative quantities of different aromatic compounds formed when several possibilities for ring closure were present. The good agreement between the calculated and observed total aromatic yield on the one hand and the simultaneous lack of agreement in the distribution of the products on the other, suggested that in certain cases isomerization must occur, a view strengthened by the reported formation of 2.5-dimethyl phenol from *n*-octyl alcohol (Komarewsky & Coley 1941).

Green & Nash (1941) have reported the formation of aromatic compounds by the passage of 2.2.4-trimethyl pentane over an aromatizing molybdenum oxide catalyst at 550° C, but Hoog, Verheus & Zuiderweg (1939) failed to obtain aromatization using a chromium oxide catalyst at

475° C. The apparent discrepancy will be shown to be due to choice of working temperature.

Hoog *et al.* (1939) suggested that secondary carbon atoms preferentially participate in ring closure. We have found evidence that certain modes of adsorption and reaction are prohibited.

The rules governing prohibition of adsorption, together with the mechanism of the first type of isomerism which is responsible for the formation of *m*- and *p*-xylene from *n*-octane, are examined in the first section of the present paper. In the second section we will deal with the products obtained by the passage over aromatizing catalysts of hydrocarbons which, without isomerization, cannot yield aromatic hydrocarbons.

EXPERIMENTAL

Chromium oxide on alumina was employed as catalyst in the form of granules *c.* 5 mm. diameter, and the hydrocarbon was injected by means of an electrolytic doser; either nitrogen or hydrogen could be used as carrier gas. In general *c.* 1 c.c. of liquid hydrocarbon was passed over 20 g. of an active catalyst.

The products were collected in a trap immersed in liquid air and the overall conversion to aromatics was determined by refractive index measurements. The value so obtained served as a basis for the choice of dilution of an aliquot portion with spectroscopically pure cyclohexane to make, in general, 1 c.c. of a solution of 1 % aromatic content. By means of a Hilger medium quartz spectrometer the ultra-violet absorption of various thicknesses of solution were compared with synthetic mixtures of known composition, and analysis was thus effected by changing the composition of the synthetic mixture until a match was obtained.

In some cases the presence of unsaturated hydrocarbons interfered with the accurate determination of the composition. In all such cases, except for the products from *n*-nonane or *n*-propyl benzene, most of the undesirable material could be removed by shaking the cyclohexane solution for 20 hr. in a sealed tube with a saturated aqueous solution of mercuric acetate. Blank experiments revealed that this treatment did not remove ordinary aromatic hydrocarbons. In this manner, 0.01 c.c. of aromatic hydrocarbon can be analysed.

The accuracy obtainable varies with the aromatic under consideration and also with the other components of the mixture. In general, the error increases in the order *p*-xylene, *m*-xylene, monosubstituted benzene, and *o*-xylene, the latter being the most difficult to determine. In the case of

a single aromatic component the error is not large, as is shown by the following example where 19 % by volume of toluene was found in the product from *n*-heptane by spectroscopic analysis as compared with 18 % by refractive index measurements. Analysis of a binary mixture consisting of *o*-xylene and benzene gave, in two independent determinations, the values 22.9, 24.0 % for *o*-xylene and 10.4, 11.4 % for benzene. In mixtures of more than two components the errors are greater and are more difficult to assess, but for a mixture such as the product from *n*-octane (table 1) the values are probably of the order 33 ± 3 % for monosubstituted derivative, 33 ± 5 % for *o*-xylene, 27 ± 2 % for *m*-xylene and 7 ± 1 % for *p*-xylene. In all cases the total aromatic found by spectroscopy agreed well with the observed refractive index.

SECTION I

In table 1 are shown the results of passage of various hydrocarbons over an alumina-supported chromium oxide catalyst at 475° C. The samples of *n*-heptane, *n*-octane, *n*-nonane and *n*-decane-decene mixture were prepared from a Fischer-Tropsch hydrocarbon-synthesis product.

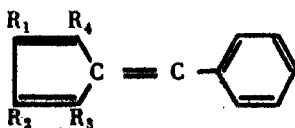
The presence of *m*- and *p*-xylene in the product from *n*-octane shows that isomerism must occur with this hydrocarbon. The only other case of the detection of an unexpected isomer from the starting materials shown in table 1 is that of *p*-xylene formed from 2,3-dimethylhexane.

Whilst 3-methyl heptane and *n*-nonane yield the expected products, the proportions are different from those anticipated from purely statistical considerations. The presence of a small quantity of indene in the product from *n*-nonane was detected by the formation of the condensation product with benzaldehyde (Thiele 1900). It is of interest to note that a solution of the benzaldehyde condensation compounds from *n*-nonane which matches in intensity in the visible a solution of the indene derivative, actually only contains about half as much indene compound when comparison is based upon the ultra-violet spectrum. The product from *n*-nonane probably contains substituted cyclopentadiene derivatives which condense with benzaldehyde to give yellow compounds of the type shown in figure 1, where R_1 , R_2 , R_3 or R_4 may be alkyl groups.

Passage of *n*-propyl benzene over an aromatizing catalyst yields a product containing β -methyl styrene and in addition a quantity of indene, which shows that subsidiary five carbon rings can be formed. Here, as was to be anticipated, the only yellow compound formed with benzaldehyde was the indene condensation product, and compounds of the type shown in figure 1 were absent.

TABLE 1. PROPERTIES OF ALIPHATIC HYDROCARBONS

hydrocarbon	η_D^{20}	$(F-C) \times 10^4$	d_4^{20}	product	% yield
heptanes: <i>n</i> -heptane	1.3882	68	0.6845	toluene	100
octanes: <i>n</i> -octane	1.3972	72	0.7032	monosubstituted benzene (mainly ethyl benzene, but contained 20 % toluene)	33
				<i>o</i> -xylene	33
				<i>m</i> -xylene	27
				<i>p</i> -xylene	7
3-methyl heptane	1.3987	72	—	ethyl benzene	15
				<i>o</i> -xylene	25
				<i>p</i> -xylene	60
3-ethyl hexane	1.4072	75	—	ethyl benzene	100
2,3-dimethyl hexane	1.4020	73	0.7125	<i>o</i> -xylene	90
				<i>p</i> -xylene	10
2-methyl heptane	1.3955	72	0.6976	<i>m</i> -xylene	100
nonanes: <i>n</i> -nonane	1.4042	71	—	<i>n</i> -propyl benzene	25-29
				<i>o</i> -methyl ethyl benzene; in addition contained trace of indene	75-71
decanes: <i>n</i> -decane-decene mixture (38 % olefine)	1.4172	80	0.7371	naphthalene	one of several components
<i>n</i> -tetradecane, mol. wt. 200	1.4295	79	—	anthracene and phenanthrene	two of many components of product; mean mol. wt. = 135

FIGURE 1. Benzaldehyde condensation product formed with material from *n*-nonane.

Naphthalene was formed when *n*-butyl benzene was passed over an aromatizing catalyst, which reconfirms the observation of Moldavski & Kamuscher (1936).

Cracking reactions are only important with paraffins higher than C_8 , for example, *n*-tetradecane (table 1), which agrees with the findings of Hoog *et al.* (1939). We will now discuss the mechanism by which *m*- and

p-xylene are formed from *n*-octane; it is important to decide whether the isomerism takes place before, during, or after ring closure. All the following hydrocarbons, *o*- and *p*-xylene, ethyl benzene, mesitylene, *trans*-1.2 and *trans*-1.4-dimethyl cyclohexane, ethyl cyclohexane and *trans*-1.3.5-trimethyl cyclohexane, failed to undergo isomerization when passed over an active catalyst, although under these conditions the naphthenes were dehydrogenated and both ethyl benzene and ethyl cyclohexane yielded some styrene. We conclude that the isomerization does not occur after ring closure.

Since the products obtained by aromatizing *n*-octane and octane-octene mixtures on two different catalysts and at two different temperatures under the conditions shown in table 2 were identical, we must conclude that isomerization takes place during the act of ring closure.

TABLE 2

hydrocarbon	catalyst	temp. °C	% aromatization
<i>n</i> -octane	Cr ₂ O ₃ on Al ₂ O ₃	475	56
"	10 % alumino vanado molybdate on Al ₂ O ₃	475	20
"	" "	550	70
octane-octene (50:50)	Cr ₂ O ₃ on Al ₂ O ₃	550	100

In figure 2 are shown the structures involved for the three octanes which can yield ethyl benzene on the simple theory of two-point adsorption of the hydrocarbon prior to ring closure, while in table 3 the values calculated from this theory are compared with the observed values.

TABLE 3. OBSERVED AND CALCULATED PRODUCTS

hydrocarbon	products	observed	calculated
<i>n</i> -octane	monosubstituted benzenes	33	66
	<i>o</i> -xylene	33	33
	<i>m</i> -xylene	27	—
	<i>p</i> -xylene	7	—
3-methyl heptane	monosubstituted benzenes	15	33
	<i>m</i> -xylene	25	33
	<i>p</i> -xylene	60	33
3-ethyl hexane	ethyl benzene	100	100

We note that for *n*-octane both *m*- and *p*-xylene have been formed at the expense of ethyl benzene so that either structure I (i) or I (iii), figure 2, undergoes isomerism, but since neither II (iv), II (v), III (i) nor III (ii)

yields *m*-xylene it seems probable that it is structure I (iii) which undergoes isomerization, since this is the only one in which an ethyl group is attached to a carbon atom which is itself undergoing reaction and is at the same time attached to the catalyst. When we consider 3-methyl heptane we shall see that structure II (v), in which an ethyl group is attached to a chemisorbed carbon atom, also shows anomalous behaviour.

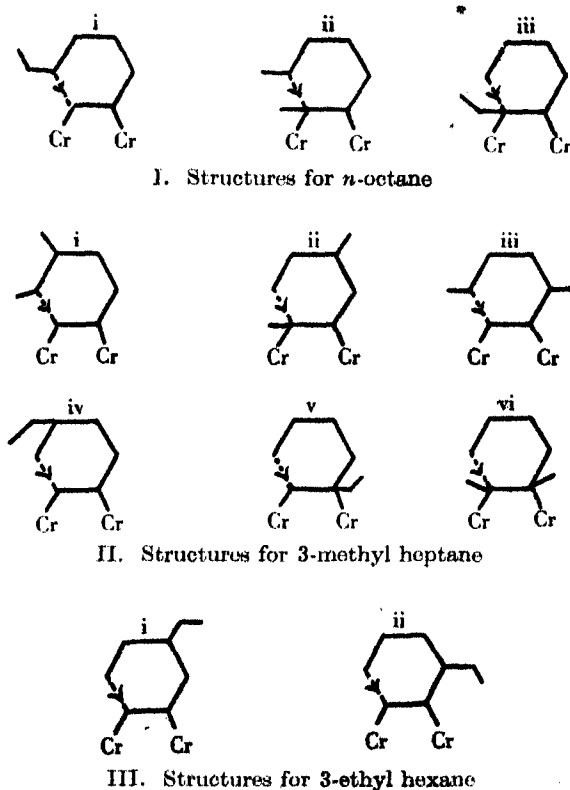


FIGURE 2

Whilst the following argument can best be understood with the aid of solid models, yet the planar diagrams (figure 3) illustrate the mechanism of the isomerization which occurs with *n*-octane. We have already advanced arguments to support the view that the changes take place concurrently with ring closure; at the moment when carbon atom 1 is linking with carbon atom 6, atom 8 may combine with 2 or 3, yielding a subsidiary five-carbon atom ring in case of *G* or a six ring in case of *H*.

Whilst the transannular ring can be accommodated for cyclohexane, dehydrogenation of the ring to an aromatic structure renders planar the six carbon atoms of the ring under which condition the ring can no longer be bridged by an ethyl group. The subsidiary ring is thus broken (dotted line) to yield *m*- or *p*-xylene. In support of this mechanism we may note that Linstead, Michaelis & Thomas (1940) report that the dehydrogenation of pinene over platinized charcoal yields *p*-cymene (see figure 4).

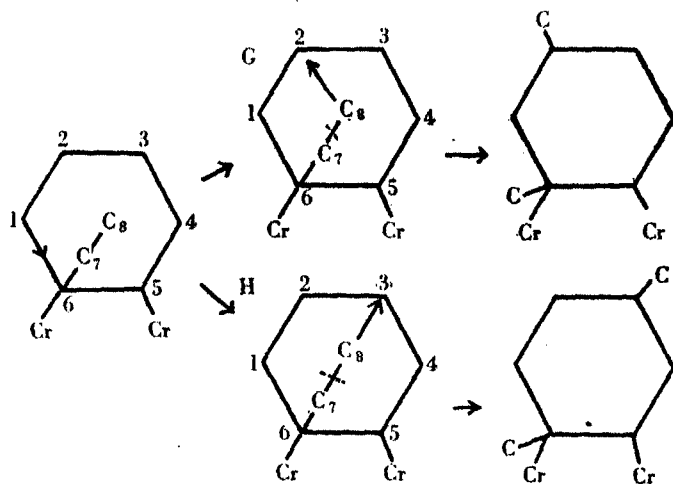


FIGURE 3. Isomerization during cyclization of *n*-octane.

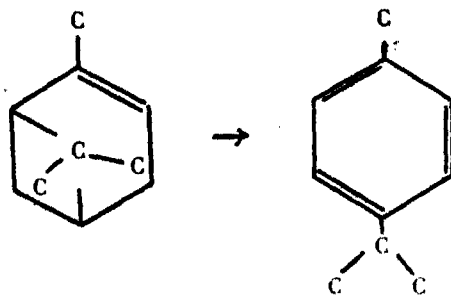


FIGURE 4. Dehydrogenation of pinene.

Moreover, this mechanism is confirmed by the calculation of the relative amounts of *m*- and *p*-xylene. Plans of the models of the boat and chair forms of cyclohexane are shown in figure 5. In these diagrams triangles represent carbon atoms, and circles or segments of circles represent hydrogen atoms; attachment of the hydrocarbon to the catalyst surface is by

two adjacent bonds on the cyclohexane ring. In the chair form both faces of the cyclohexane ring are similar, (a) and (b), and both have three hydrogen atoms in the plane and six in the periphery. In the boat form there are four hydrogen atoms on one face (c) and two on the other (d). By means of scale models it can be shown that if one hydrogen atom (1) in (a), (b) or (c) be replaced by an ethyl group, then these three structures will give *m*-xylene by the reaction shown in figure 3. For example, if the ethyl group be on carbon 1 (figure 5 (c)), then the link will be with carbon 3. Configuration (d) will, however, yield *p*-xylene.

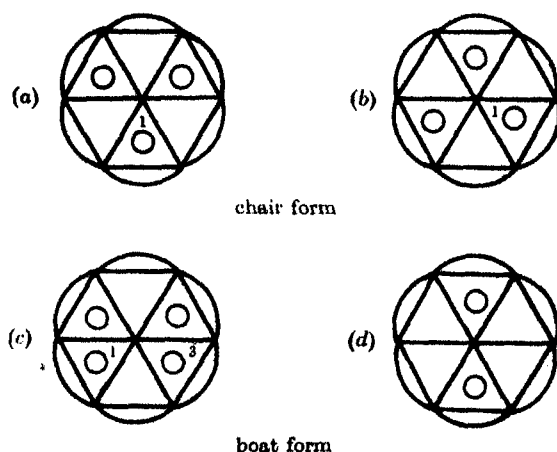


FIGURE 5. Diagrammatic representation of solid models of two cyclohexane forms.

If we assume that at the moment of ring closure the probabilities that the chain exists in the boat or chair form are equal, then calling the probability it exists in configuration (a) at this moment p , then the probability that it will be (b) will likewise be p , then the total probability it will be (c) plus (d) will be $2p$. There are four ways in which it may be (c) and only two in which it may be (d), so the probability for (c) is $\frac{4}{3} 2p = \frac{8}{3}p$, and for (d) is $\frac{2}{3} 2p = \frac{4}{3}p$. Hence the final product will result as in table 4.

TABLE 4

product	structure contributing	probability
<i>m</i> -xylene	figure 5 (a)	p
	(b)	p
	(c)	$\frac{8}{3}p$
		total $\frac{14}{3}p$
<i>p</i> -xylene	figure 5 (d)	$\frac{4}{3}p$

i.e. ratio *meta*/*para* = 5/1.

In table 5 are shown the values calculated on this basis as compared with observed. This mechanism does not affect the arguments for overall conversion previously published (Herington & Rideal 1945).

TABLE 5. PRODUCTS FROM *n*-OCTANE

product	% observed	% calculated	overall rate	
			observed	calculated
monosubstituted benzene	33	33	46	42
<i>o</i> -xylene	33	33	—	—
<i>m</i> -xylene	27	28	—	—
<i>p</i> -xylene	7	6	—	—

3-methyl heptane exhibits a different type of anomalous behaviour, for although the product does not contain unexpected aromatics, yet the distribution is very different from the values calculated on the assumption of equal probabilities for the different configurations (table 3). One each of the structures (figure 2) which should yield ethyl benzene and *o*-xylene appear to be suppressed. Structures II (v) and II (vi) (figure 2) differ from all the other structures in this figure in so far as II (v) contains an ethyl group attached to a carbon atom adjacent to the position of formation of a new link, while at the same time the ethyl group is attached to a carbon atom which is chemisorbed. In structure II (vi) both chemisorbed carbon atoms carry methyl groups. Steric hindrance may be involved in this type of adsorption leading to suppression of these configurations, a view which receives support from similar behaviour of adsorbed chains in the five-six ring isomerization (§ 2, see p. 379).

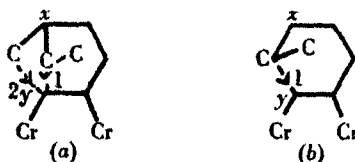


FIGURE 6. Structures for 3-methyl heptane (a) which yield structures II (i) and II (iv) (figure 2) as compared with *n*-heptane (b).

Structures II (iv) and II (i) in figure 2 originate from adsorption on the same pair of carbon atoms. To obtain the calculated values shown in table 3 it was assumed that each of these structures made the same contribution as each of the other four. This is equivalent to assuming that the probability of formation of links (1) and (2), figure 6 (a), are equal and both have the same value as for the formation of the link in *n*-heptane, figure 6 (b), i.e. the velocity of ring closure for the mode of adsorption shown

in figure 6 (a) is twice that for *n*-heptane in figure 6 (b). If, however, the slow step in the ring closure is the time taken for the chain from C_x to C_y (figure 6), to obtain the correct configuration, then once C_x is in the correct position, link 1 will probably rapidly follow. If this be the case, then the probability of C_y reacting in figure 6 (a) is not twice that of C_y in figure 6 (b). The value for the probability of C_y reacting in figure 6 (a) as compared with C_y for figure 6 (b) can be computed from the analytical data and hence the overall reaction rate may be found.

If the probability that the hydrocarbon reacts so as to give ethyl benzene as in figure 6 (a) link 2 be p_e , then the probability it reacts to give *o*-xylene (figure 6a link 1) will likewise be p_e . Let the probabilities of reaction for structures II (ii) and II (iii) (figure 2) be P . Then remembering that structures II (v) and II (vi) do not contribute, we have

$$\text{Ratio } p\text{-xylene}/(\text{Ethyl benzene} + o\text{-xylene}) = \frac{2P}{2P_e}.$$

From experiment this ratio = 60/40, i.e. $P_e = \frac{2}{3}P$.

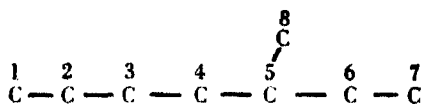


FIGURE 7. 3-methyl heptane.

Consider the adsorption of 3-methyl heptane; adsorption of the pairs of carbon atoms 4-5, 5-6 and 5-8 (figure 7) will be blocked because of steric factors, but the pairs 1-2, 2-3, 3-4 and 6-7 can be chemisorbed. Hence, the probability of total reaction = $\frac{\frac{4}{3}P + 2P}{4} = 0.83P$. This value is very near that given in the first paper (Herington & Rideal 1942), so that the value shown there is practically unchanged. The following results (table 6) are thus obtained:

TABLE 6

products	% overall reaction rates	
	observed	calculated
ethyl benzene	15	20
<i>o</i> -xylene	25	20
<i>p</i> -xylene	60	60

If *n*-nonane reacted in a similar manner to *n*-octane we should expect to find *m*- and *p*-methyl ethyl benzene together with 1.2.3 and 1.2.4-trimethyl benzene in the products, these having been formed by the reactions shown in figure 8.

Actually these are absent from the product. However, for *n*-nonane new possibilities arise which are not present for *n*-octane; these are shown in figure 9. If these reactions (figure 9) take place more readily than bridging the ring, then the absence of *m*- and *p*-methyl ethyl benzene, 1.2.3 and 1.2.4-trimethyl benzene is explicable. Moreover, the ratio of *o*-methyl ethyl benzene to *n*-propyl benzene should be as 3:1, since all the structures in figure 9 except (1) will yield *ortho*-substituted derivatives. By this mechanism the overall rate will be the same as that calculated from the simple theory. Since hydriodene is readily dehydrogenated to indene by passage over the catalyst we must conclude that the isomerization of *n*-nonane accompanies ring closure.

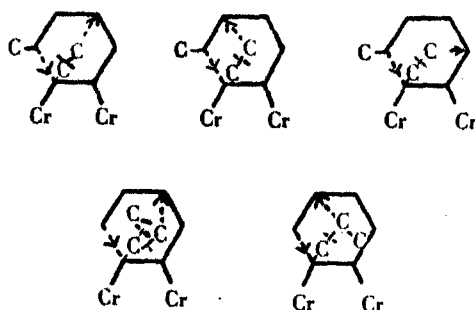


FIGURE 8. Possible isomerization of *n*-nonane by analogy with *n*-octane.

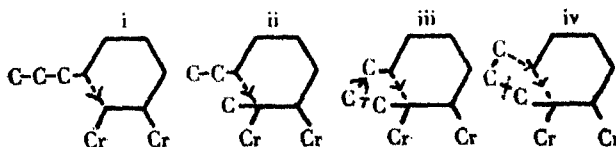


FIGURE 9. Actual reactions of *n*-nonane showing all configurations.

SECTION 2

We consider here the mechanism of aromatization of aliphatic hydrocarbons which can only yield aromatics by isomerization. The experimental procedure was identical with that described in § 1. The hydrocarbons examined and the results obtained are given in table 7.

It will be noted that all these substituted pentanes yield aromatics on passage over an aromatizing catalyst, but they differ from the hexanes and higher hydrocarbons in that the temperature required for reaction is higher. Thus 2.2.4-trimethyl pentane yields only 4 % aromatics at 475° C and 50 % at 550° C, whilst under similar conditions *n*-octane

yielded 40 and 100 % aromatics respectively. The 3.3 and 2.2-dimethyl hexane are of interest, because although direct ring closure is possible they cannot yield aromatics without a migration of a methyl group. Here again only 1.5 % aromatics are formed at 475° C, and at 550° C, 50 %. The small quantities of toluene formed (for exceptions see p. 378) from the hydrocarbons containing eight carbon atoms and the absence of benzene and dimethyl naphthalenes in the product from 2.3-dimethyl pentane are strong evidence that the mechanism of isomerization does not involve the cracking of the hydrocarbon into two separate fragments which then recombine in a different way.

TABLE 7

hydrocarbon	b.p. °C at 760 mm.	η_D^{20}	$(F-C)$ $\times 10^4$	d_4^{20}	product	yield %
2.3-dimethyl pentane	—	1.3925	71	0.6949	toluene	100
2.2.3-trimethyl pentane	110.0	1.4029	73	0.7156	<i>m</i> -xylene	100
2.2.4-trimethyl pentane	99.4	1.3918	73	0.6916	<i>p</i> -xylene	100
2.3.3-trimethyl pentane	114.7	1.4079	74	0.7258	<i>m</i> -xylene	100
2.3.4-trimethyl pentane	113.8	1.4045	75	0.7182	<i>o</i> -xylene	80
					<i>p</i> -xylene	40
2-methyl-3-ethyl pentane	115.7	1.4046	75	0.7191	ethyl benzene	12.5
					toluene	12.5
					<i>o</i> -xylene	25.0
					<i>m</i> -xylene	50.0
3-methyl-3-ethyl pentane	118.5	1.4079	73	0.7272	ethyl benzene	24
					toluene	24
					<i>o</i> -xylene	48
					<i>m</i> -xylene	4
3.3-dimethyl hexane	111.7	1.4007	73	0.7103	ethyl benzene	10
					toluene	10
					<i>o</i> -xylene	40
					<i>m</i> -xylene	20
					<i>p</i> -xylene	20
2.2-dimethyl hexane	107.1	1.3940	74	0.6954	<i>m</i> -xylene	100

In view of the great ease with which the reaction cyclohexane \rightleftharpoons methyl cyclopentane proceeds under certain circumstances, a five-six ring isomerization for the mechanism of aromatization of these hydrocarbons seems probable. The general type of reaction can be represented by considering the case of 3-methyl pentane (figure 10). A study of the action of the catalyst on the 2.3-dimethyl butane and on 2.2.3.3-tetramethyl butane confirms this view, for the former gives rise to 2.3-dimethyl butadiene but no benzene, whilst the latter passes through unchanged. No aromatics are formed since no five-six ring isomerization is possible.

Further evidence in support of this mechanism is furnished by the observation that both ethyl cyclopentane and *trans*-1,2-dimethyl cyclopentane give toluene when passed over the aromatizing catalyst at 550° C. The same catalyst yielded a product containing 30 % toluene from 2,3-dimethyl pentane, 20 % from ethyl cyclopentane, and 27 % from *trans*-1,2-dimethyl cyclopentane under identical conditions. These figures show that the rate of aromatization from cyclopentane derivatives is not greater but rather smaller than from the substituted pentanes,

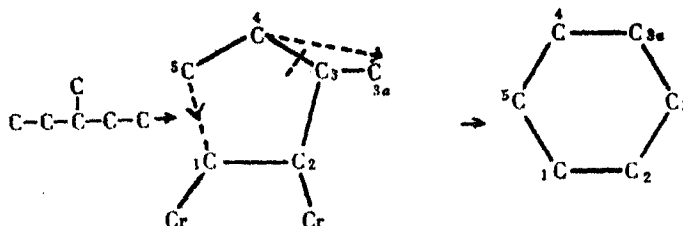


FIGURE 10. Aromatization of 3-methyl pentane.

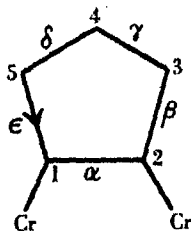


FIGURE 11. Designation of bonds in five-six ring isomerism.

suggesting that the slow step in the overall conversion of these substituted pentanes is the rupture of the five ring, and that conditions following the formation of a five ring from a pentane are advantageous for subsequent fission. The rates of aromatization of all the hydrocarbons cited in table 7 are practically the same, the only step identical in all being the fission of the five ring. Analysis of the products of aromatization permits us to make certain generalizations concerning the five-six ring reaction.

If we consider the pentane chain chemisorbed at atoms C_1 and C_2 with the ring closing from C_5 (figure 11), we find that, in general, the bond γ is broken unless C_3 carries an ethyl group, or C_2 and C_3 both carry a methyl group, in which case it is the bond β which is broken.

Certain rules can be formulated for the subsequent closure to a six ring. Thus if C_3 or C_4 is a quaternary carbon atom, the link is remade to one

of the carbon atoms attached to this quaternary atom. If there be no quaternary carbon atom, then the links with either of the two groups adjacent to the bond broken are equally likely, as shown in figure 12. Finally, it appears that if ring closure should produce ethyl benzene, half of this actually appears in the product as toluene.

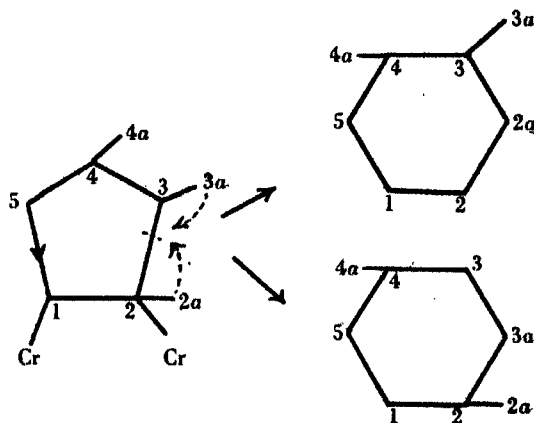


FIGURE 12. Remaking bonds in case of 2,3,4-trimethyl pentane.

		percentages	
		observed	calculated
2,3-dimethyl pentane	toluene	100	100
2,2,3-trimethyl pentane	<i>m</i> -xylene	100	100
2,2,4-trimethyl pentane	<i>p</i> -xylene	100	100
2,3,3-trimethyl pentane	<i>m</i> -xylene	100	100
2,3,4-trimethyl pentane	<i>o</i> -xylene	60	50
	<i>p</i> -xylene	40	50
2-methyl-3-ethyl pentane	ethyl benzene	12.5	12.5
	toluene	12.5	12.5
	<i>o</i> -xylene	25.0	25.0
	<i>m</i> -xylene	50.0	50.0
3-methyl-3-ethyl pentane	ethyl benzene	24	25
	toluene	24	25
	<i>o</i> -xylene	48	50
	<i>p</i> -xylene	4	—
3,3-dimethyl hexane	ethyl benzene	10	10
	toluene	10	10
	<i>o</i> -xylene	40	40
	<i>m</i> -xylene	20	20
	<i>p</i> -xylene	20	20
2,2-dimethyl hexane	<i>m</i> -xylene	100	100

In table 8 are shown the observed yields compared with the calculated values obtained by the application of the above rules in conjunction with the same assumptions for the formation of the five ring as were used for the six ring (Herington & Rideal 1945). The rules which determine whether

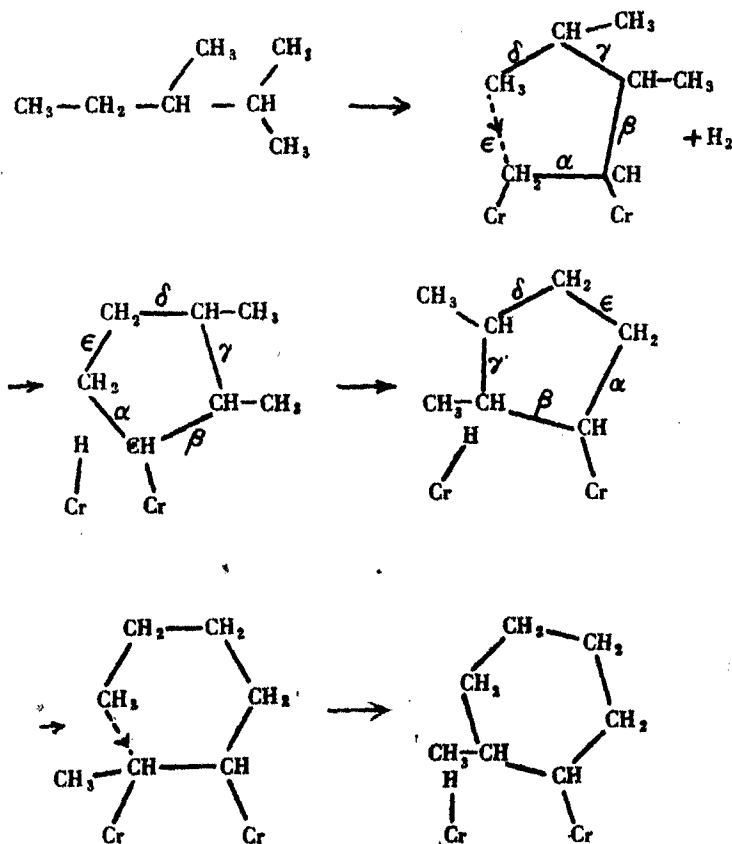


FIGURE 13. Aromatization of 2,3-dimethyl pentane, five-six ring isomerism, bond γ broken.

bond β or bond γ be broken in figure 11 are related to the conditions which inhibit modes of adsorption such as II (v) and II (vi) in figure 2. The sequence of steps leading to the formation of toluene from 2,3-dimethyl pentane can be depicted as in figure 13. Thus bond γ is broken by a mechanism which resembles that proposed for double-bond migration in olefines (Twigg 1939, 1941; Twigg & Rideal 1939, 1940). In the present instance the rupture of the link is effected by the addition of a hydrogen

atom which has been liberated by the formation of the five-carbon ring. After formation, the five ring partially rotates and is adsorbed following attack by the chemisorbed hydrogen atom. This explains why the speed of aromatization of substituted pentanes is greater than for substituted

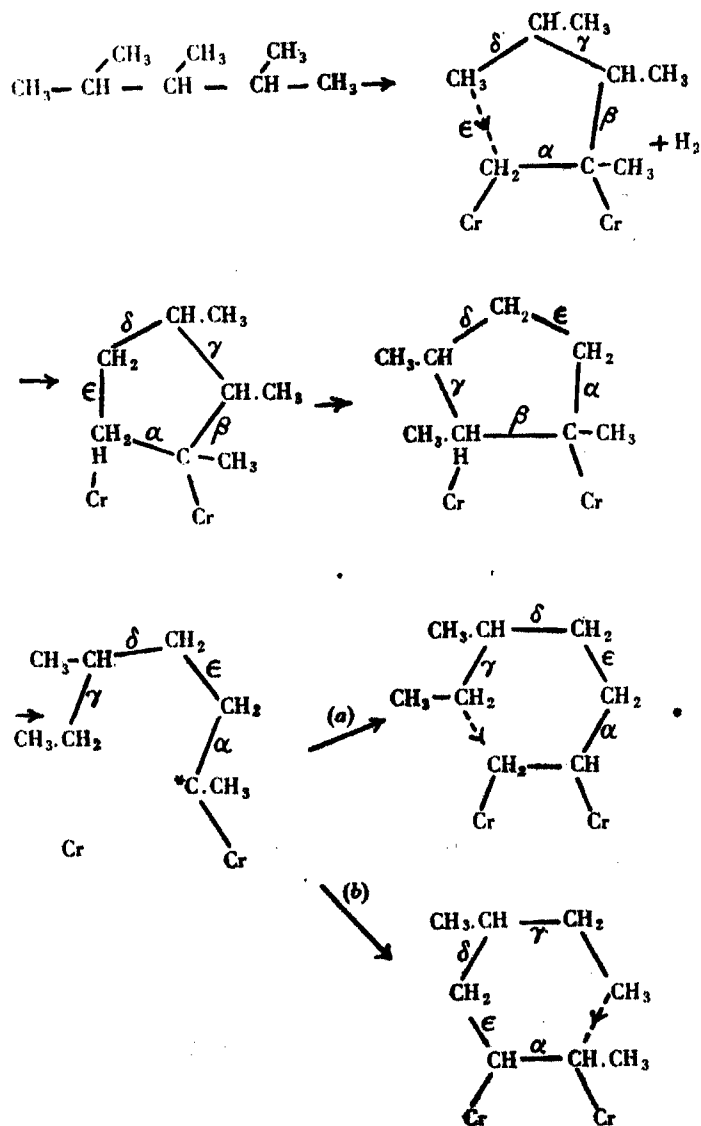


FIGURE 14. Aromatization of 2,3,4-trimethyl pentane, five-six ring isomerism, bond β broken.

cyclopentanes, although the latter would appear to involve fewer steps. The pentanes automatically liberate a hydrogen atom in the correct position for the cracking reaction, while the cyclopentanes do not.

If, however, C_2 carries an ethyl group, or C_2 and C_3 both carry a methyl group, bond β (figure 11) cannot be adsorbed and it breaks instead. The mechanism in this case is different and proceeds through a radical as shown in figure 14.

Here again a bond is broken by hydrogen attack, but this is followed by adsorption of either the methyl group attached to carbon C^* , or by adsorption of the adjacent methylene group. We may note that the adsorption in the last stage, (a) figure 14, only requires the rotation of a CH_3 group, while (b) entails the whole molecule, and experimentally the ratio a/b is not unity but is as 60/40.

We wish to express our thanks to Professor R. G. W. Norrish for facilities in the use of the ultra-violet spectrometer, to Dr W. C. Price for advice in connexion with analysis by means of the spectra, and to Dr G. B. B. M. Sutherland and Dr E. E. Turner for specimens of pure hydrocarbons.

This work, which forms part of the programme of the Fuel Research Board of the Department of Scientific and Industrial Research, has been carried out in the Department of Colloid Science, Cambridge. The account is printed by permission of the Director of Fuel Research. The illustrations are Crown copyright and are reproduced by permission of the Controller of H.M. Stationery Office.

REFERENCES

- Green, S. W. & Nash, A. W. 1941 *Chem. and Ind.* **60**, 801.
Herington, E. F. G. & Rideal, E. K. 1945 *Proc. Roy. Soc. A*, **184**, 434.
Hoog, H., Verheus, J. & Zuiderweg, F. J. 1939 *Trans. Faraday Soc.* **35**, 993.
Komarewsky, V. I. & Coley, J. R. 1941 *J. Amer. Chem. Soc.* **63**, 700.
Linstead, R. P., Michaelis, K. O. A. & Thomas, S. L. S. 1940 *J. Chem. Soc.* p. 1139.
Moldavski, B. & Kamuscher, H. 1936 *C.R. Acad. Sci. U.R.S.S.* **1**, 355.
Thiele, J. 1900 *Ber. dtsch. chem. Ges.* **33**, 3395.
Twigg, G. H. 1939 *Trans. Faraday Soc.* **35**, 934.
Twigg, G. H. 1941 *Proc. Roy. Soc. A*, **178**, 106.
Twigg, G. H. & Rideal, E. K. 1939 *Proc. Roy. Soc. A*, **171**, 55.
Twigg, G. H. & Rideal, E. K. 1940 *Trans. Faraday Soc.* **36**, 533.

The age constitution and the future fellowship of the Royal Society

BY LEON SOLOMON, M.A., PH.D.

(Communicated by Professor P. M. S. Blackett, F.R.S.—Received 23 March 1945)

The ordinary fellowship of the Royal Society has been steadily ageing over the period investigated, 1899–1943 inclusive. This has been due in part to the natural growth of the Society towards actuarial stability, and in part to the increasing age of Fellows at election. The latter trend has been in progress since 1848, and has not been reversed by the increases, in 1931 and 1938, of the total number of annual elections.

The mortality rates experienced during 1899–1943 have been abstracted and differences noted between them and those of various standard tables. The crude results were graduated by fitting a Makeham curve of the form

$$q_x = A + Bc^x,$$

where q_x is the probability that a Fellow of exact age x will die within one year. A , B , c are constants.

There is some evidence that the mortality rates of older Fellows (age 75 and greater) were heavier in the latter part of the period investigated than in the former part.

The graduated mortality rates were used to forecast the future size and age constitution of the Royal Society. The results are given for several different assumptions regarding the number and ages of future successful candidates for the Fellowship.

INTRODUCTION

The periodicals of the Royal Society contain few contributions on the life statistics of the Fellowship. Two papers have appeared in the *Proceedings* (Strachey 1892; Schuster 1922), while a note on the ages at election to the Royal Society was published in *Notes and Records* (Hill 1939). The present paper attempts to bring the relevant statistics up to date, to analyse the trends of recent years, and to hazard some estimates concerning the future.

Strachey investigated 'the probable effect of the limitation of the number of ordinary Fellows into the Royal Society to 15 in each year on the eventual total numbers of Fellows', and concluded that the latter would be about 415, and certainly not more than 425. In fact, the total Fellowship produced by the annual election of 15 Fellows averaged 437 during 1900–30. In the course of his paper, Strachey estimated that the mortality of Fellows had been some 5 % lighter than Dr Farr's Life Table, which was based on the vital statistics for the male population of England and Wales between 1837 and 1854. As it has happened, the mortality has been much more favourable than Strachey expected. This accounts sufficiently for the discrepancy between the estimated and actual totals of the Fellowship in its steady state.

Schuster was interested, among other topics, in the expectancy of life, at various ages, of Fellows of the Royal Society. He realized, rightly, that the mean of the completed lifetimes of deceased Fellows was an insufficient estimate, and that some allowance ought to be made for the future span of life of those who formed the

Society at the date of his investigation. 'For the Fellows who are still living we are ignorant of their ultimate life, but we find an approximate value by adding to the number (a) of years which have elapsed since election up to the end of 1922 their further expectancy (w) taken from some reliable table of life statistics.' Instead, however, of the *further* expectancy of living Fellows, Schuster apparently added the original expectancy for their age at election—which must have led, in the case of the older Fellows in particular, to a serious over-estimate of future lifetime. His conclusion, that Fellows enjoy a considerably longer expectation of life than the general population of the same age, is nevertheless valid, as the following pages reveal.

In the following sections, various statistics concerning the ordinary Fellowship of the Royal Society are discussed. In this work, 'Statute 11' and 'Statute 12' Fellows (Royal personages and non-scientific persons who have benefited or may benefit science, respectively), as well as Foreign Members, are excluded.

THE AGE CONSTITUTION OF THE FELLOWSHIP

In 1848 the number of elections into the Royal Society was limited to 15 per annum, and by the early years of this century the ordinary Fellowship had virtually reached a steady state. Thereafter, until 1930, the total number fluctuated around the mean value of 437 Fellows. In 1931 and in each subsequent year until 1937 (inclusive), 17 Fellows were elected each year, instead of 15. A further increase in the number of elections to 20 annually occurred in 1938, and this figure has been adhered to up to the present time. In consequence of these changes, the total fellowship increased slowly, and reached 455 on 1 January 1944. During the preceding five years, the rate of growth has averaged about 2 Fellows a year.

Even before 1930, however, the Society was not truly in equilibrium. It was, on the contrary, ageing slowly. The salient features of this trend are summarized in table 1, which has been compiled from the age distribution of Fellows for the years 1899, 1904, 1909, ..., 1944. No important tendencies are distorted by grouping these data into only three periods, 1899–1914, 1915–30 and 1931–44. The points of division were selected to separate, first, the pre-war period from the subsequent epoch, and, secondly, the years in which only 15 annual elections were made, from those in which this number was raised to 17 and thereafter to 20.

The median age is such that half the Fellows are of greater, and half of lesser, age. One-quarter of the Fellows are younger, and three-quarters are older than the 'lower quartile' age; similarly, three-quarters are younger, and one-quarter older than the 'upper quartile' age.

The ageing tendency is apparent both from the proportions of Fellows in the various age groups of table 1, and from the slight, but probably significant, upward trend in the median and quartile ages. It is noteworthy that the percentage of Fellows of 'intermediate' age (45–54) has remained virtually constant over 50 years; while the percentage of younger (under 45) and older (55 and over) Fellows have, respectively, diminished and increased by appreciable amounts.

TABLE 1. AVERAGE PROPORTION OF TOTAL FELLOWSHIP BETWEEN GIVEN AGES

ages	1899-1914 %	1915-1930 %	1931-1944 %
under 35	1	0.5	0.5
35-44	13	7.5	8.5
45-54	25	22	23
55-64	26	34	27
65 and over	35	36	41
	100	100	100
lower quartile age	50 years	53 years	53 years
median age	59 years	60 years	62 years
upper quartile age	69 years	69 years	71 years

Another point of interest is that the figures in the last column in particular of table 1 do not refer to a Society 'in equilibrium'. As has been remarked, the increases in 1931 and 1938 in the number of annual elections have caused the Royal Society to grow in size, and, if present conditions are maintained, it will continue both to grow and slightly to age for the next quarter century.

AGES AT ELECTION

The average age of Fellows at election has been rising steadily for the past century. The dominant reason has been the marked reduction in the number of younger Fellows elected (age 30-40). This of course is the explanation of the decline in the proportion of such Fellows in the Society as a whole, recorded in table 1. Table 2 illustrates the extent of these changes (data for the years prior to 1920 are derived from Schuster's paper).

TABLE 2. AGES OF FELLOWS AT ELECTION

period	lower quartile age years	median age years	upper quartile age years	interquar- tile range years	proportion of Fellows elected under age 35 %
1848-1872	36.5	43	51.5	15	20
1873-1897	37	44	53	16	16
1898-1922	39.5	45.5	52	12.5	10
1920-1930	40	47	53.5	13.5	10
1931-1937	42	47.5	53.5	11.5	2
1938-1944	42	48	53.5	11	4

The principal features revealed by the figures may be briefly enumerated. They are:

(a) The slow but steady increase in the median age, and the rather more marked increase in the lower quartile age.

(b) The absence of any variation in the upper quartile age, apart from minor fluctuations.

It follows that the tendency to a greater age at election was mainly due to a diminution in the number of Fellows admitted while still young—a fact strikingly confirmed by the figures of the last column. The tendency is not due to any increase in the number of successful candidates of relatively 'advanced' age (over 55).

In consequence of these tendencies, the dispersion of ages at election is now materially less than during the last century. This is illustrated in the penultimate column of table 2, in which the 'interquartile range'—or upper quartile age less lower quartile age—has been chosen as a convenient measure of dispersion. From about 15½ years in the period 1848–97, the interquartile range fell to 11 years (1931–44).

A final point of some interest concerns the effect on the ages of successful candidates of raising the number of annual elections. The increases in the latter, from 15 to 17 in 1931, and to 20 in 1938, may possibly have retarded the general tendency to elect older men, but they did not reverse it. For the recent period 1920–44, the age distributions of Fellows at election may be quoted (table 3), as supplementing the more compressed information of table 2.

TABLE 3

age at election	proportion of Fellows elected at given ages			
	1920–30	1931–37	1938–44	1931–44
	%	%	%	%
under 30	0·6	0	0	0
30–39	24·4	19	19	19
40–49	25	42	40	41
50–59	26	28	32	30
60 and over	9	11	9	10
	100	100	100	100

MORTALITY EXPERIENCE

The derivation of the rates of mortality of the Royal Society is the preliminary to any forecast of its future actuarial structure. Such vital statistics are also of interest when compared with other sets, in casting some light on the effect of social and economic status on vitality. The importance of such factors is well known. But how important they are, and how far their influence is direct, and how far indirect, one cannot yet state with precision.

The data used in the present mortality investigation of the Royal Society were:

(a) The numbers of Fellows classified according to age, on 1 January of the years 1899, 1904, 1909, ..., 1944.

(b) The list of Fellows who died between 1 January 1899 and 31 December 1943, tabulated according to age at, and calendar year of, death.

The procedure adopted was the standard 'census' method, so called because, following on the periodic national censuses, it has been employed to ascertain the mortality experience of the country. The details have been sufficiently described elsewhere (Elderton & Fippard 1935; Hardy 1909).

Table 4 sets out the results. In the first column are quoted the 'years exposed to risk' at various quinquennial ages (x). This quantity is the total number of years lived by Fellows, over the period of the investigation, while between exact age x and exact age $(x+1)$. Fellows who died between these ages are reckoned as having contributed a complete year to the 'exposure'. (For the purpose of this procedure see Elderton & Fippard (1935).) The second and third columns contain respectively the estimated and 'smoothed' rates of mortality (q_x)—that is, the probability that a Fellow, of exact age x , will die before attaining age $(x+1)$. It follows that the number of deaths between these ages is the product of the 'exposed to risk' and q_x . The remaining columns of the table give the mortality rates from certain standard life tables, which are described below.

TABLE 4

age (x)	exposed to risk (i) years	crude q_x (ii)	smoothed q_x (iii)	q_x in some standard tables			
				E.L. 10 (iv)	E.R.R. (v)	A. 1924-9 ult. (vi)	$a(m)$ ult. (vii)
37	579	0.003	0.00226	0.00474	0.00340	0.00320	—
42	1333	0.0035	0.00363	0.00639	0.00443	0.00439	0.00818
47	2070	0.0080	0.00575	0.00925	0.00621	0.00604	0.00883
52	2648	0.0054	0.00901	0.01295	0.00813	0.00906	0.01122
57	2900	0.0104	0.01407	0.01890	0.01252	0.01450	0.01651
62	2870	0.0231	0.02189	0.02875	0.02024	0.02394	0.02422
67	2610	0.0359	0.03399	0.04568	0.03401	0.03908	0.03550
72	2147	0.0480	0.05270	0.07246	0.05551	0.06473	0.05217
77	1565	0.0877	0.08164	0.11325	0.09142	0.10093	0.08378
82	844	0.1311	0.12641	0.16927	0.14830	0.15066	0.12697
87	352	0.203	0.19563	0.24078	—	0.21315	0.18409
92	64	0.308	0.30272	0.31998	—	0.28700	0.26276
97	7	0.48	0.46834	0.41720	—	0.37231	0.36664

The method of smoothing the crude results is discussed later. In the meantime, we note that the graduated probabilities of column (iii) are reasonably close to the observed values, and that the differences between the two sets alternate fairly frequently from positive to negative, so that there is no evidence of systematic bias. The agreement is in fact as good as can reasonably be expected (see below).

'English Life Table Number 10—Males' (Registrar-General 1936) (abbreviated to E.L. 10 in table 4) was constructed from the mortality experience of the male population of England and Wales during 1930-2 and is the latest published national life table. Clearly, the rates of mortality experienced by Fellows of the Royal Society between 1899 and 1944 were very materially lighter at all ages (except the last, 97) than those of the population as a whole in 1930-2. In view of the wide difference in average social and economic status of the two bodies of men, the divergence is, perhaps, not surprising.

In the *Registrar-General's Report* (1936) the mortality during 1930-2 was also examined for eleven geographical regions of England and Wales, each region being

subdivided into county boroughs, urban districts, and rural districts. It was found that the lightest mortality of all (at practically every age) had been sustained in the rural districts of the east region, which comprised Cambridgeshire, Isle of Ely, Huntingdonshire, Lincolnshire (parts of Holland, Kesteven, and Lindsey), Norfolk, Rutlandshire, east Suffolk, and west Suffolk. The relevant rates, for males, are quoted in table 4, column (v). It is interesting that, at ages where comparison is possible, they do not materially differ from the Royal Society figures. These rates, it may be emphasized, refer to a geographical, and to a certain extent (being restricted to rural districts) an occupational subgroup of the population, but no social classification is involved. The fact that men in these districts should have been subject in 1930-2 to mortality more or less the same as that of Fellows of the Royal Society in the last half-century, is probably no more than an accident.

Column (iv) of table 4 provides a comparison of a different nature. The A. 1924-9 (ultimate) Table (1935) was based on the experience of male policyholders in most of the ordinary British life assurance offices, during the years indicated. The restriction to ordinary, as opposed to industrial, offices excluded working-class 'lives', to a large extent, from the survey. The term 'ultimate' may require clarification. On effecting an assurance contract, most policyholders had either passed a medical examination, or had given satisfactory evidence of good health. They thus formed a medically selected body of men, and it was found that they enjoyed for some time much lighter mortality than others, of the same age, whose policies had already been in force for some years. The effect of the initial medical selection wore off eventually, and the rates of mortality, after its disappearance, were termed 'ultimate'. Comparing these ultimate rates with those of the Royal Society, we conclude that the latter experienced lighter mortality than non-working class policyholders, no longer medically selected, of British life assurance offices.

The final comparison of table 4 is with another life office standard. The *a(m)* (ultimate) table was constructed from the statistics of males between 1900 and 1920 who were receiving annuities from British assurance offices (Elderton & Oakley 1924). Again, a systematic divergence from Royal Society results is apparent. At the younger and middle ages (72 and less) Fellows experienced the lighter rates; but at more advanced years they suffered more heavily than the annuitants.

Merely to compare a standard *rate* of mortality with that estimated in some new investigation ignores the weight attaching to the latter. A better procedure is to compare, for individual ages or age groups, the actual number of deaths, with the number to be expected on any proposed standard basis. The agreement between the two may then be examined statistically in a number of ways, of which the χ^2 test of Karl Pearson is one. The validity of this test, as applied to mortality data, has recently been critically examined, and confirmed by Seal (1941).

Table 5 compares the actual deaths with those expected from the above-described standards. The χ^2 values are obtained from contributions at quinquennial ages (37, 42, etc.). The quantity *P* represents the probability that the observed, or a greater value of χ^2 , could have arisen merely from the effects of random sampling.

TABLE 5

basis	number of deaths expected	actual deaths —expected deaths at same ages	χ^2	degrees of freedom in χ^2	P
actual experience	673 (ages 37 and over)	—	—	—	—
smoothed data	673 (ages 37 and over)	0	10.6	9	0.30
E.L. 10	905 (ages 37 and over)	— 232	58.3	11	negligible
E.R.R.	613 (ages 37–82)	— 34	10.1	9	0.33
A. 1924–9 ult.	777 (ages 37 and over)	— 104	23.9	11	0.013
$a(m)$ ult.	702 (ages 42 and over)	— 31	22.8	11	0.035

In the search for a standard to which the experience of the Royal Society would conform sufficiently closely the χ^2 test would eliminate English Life Table 10 without any doubt, and, with negligible risk of error, the A. 1924–9 and $a(m)$ tables. The smoothed rates and the east region rural districts experience pass this test, and, indeed, the latter might reasonably have been accepted. The main disadvantages are that it does not cover the whole range of ages required and that an actual life table of this experience, as distinct from the mortality rates q_x , has not been published.

The graduation formula which, after some trials, was finally adopted was

$$q_x = A + Bc^x,$$

where A , B , and c are constants, and x denotes the age. The criterion for determining the best values of the parameters was that of least squares, subject to the proviso that the total 'expected deaths' should equal the total number of actual deaths. In its above-quoted form, the formula is inconvenient for making a 'least squares' estimate of c . A transformation, which the author some years ago suggested would facilitate the computation, was therefore used. If the expression for q_x is summed, from the youngest age in question—let us call this age zero—to age $(t-1)$, there results

$$\sum_0^{t-1} q_x = At + \frac{q_t}{c-1} - \frac{A+B}{c-1},$$

from which the parameters A , $(c-1)$ and $(A+B)/(c-1)$ can be readily obtained by the customary 'least squares' procedure. The values obtained gave the formula

$$q_x = -0.00120 + 0.00250(1.09114)^{x-37}.$$

This formula was originally suggested, as an analytical description of certain mortality experiences, by Gompertz (1825), and was developed by Makeham (1860),* who later attempted to found this 'law' on physiological grounds. It is doubtful, however, if this justification were generally accepted. Nevertheless, the formula proved of wide applicability. In recent years it has tended to pass

* Strictly, Makeham suggested that the formula $A + Bc^x$ described the age variation of the force of mortality (μ_x), i.e. the rate of mortality experienced over the infinitesimal age range x to $x+dx$. For most practical purposes, it is sufficiently accurate to take $q_x = \mu_{x+1}$.

out of actuarial favour—but whether mortality in the twentieth century has ceased to conform to Makeham's Law, or whether the formula has been tested in such instances on heterogeneous data, remain moot points. Its success in describing the—actuarially speaking—small volume of Royal Society data is interesting, but is hardly of major importance in this connexion. Another feature of some interest is that, in many independent investigations, covering widely differing periods of time, classes of persons, and nations, the value of the Makeham parameter c has been nearly constant, and of the order of 1.1. Some writers have speculated whether this striking result may not imply the existence of a 'universal demographic constant'.

TABLE 6. LIFE TABLE OF THE ROYAL SOCIETY, 1899–1943 INCLUSIVE

age x	q_x	p_x	l_x	age x	q_x	p_x	l_x
32	0.00138	0.99862	100,000	67	0.03399	0.96601	70,295
3	0.00152	0.99848	99,862	8	0.03711	0.96289	67,906
4	0.00168	0.99832	99,710	9	0.04051	0.95949	65,386
5	0.00186	0.99814	99,543	70	0.04423	0.95577	62,736
6	0.00205	0.99795	99,358	1	0.04828	0.95172	59,961
37	0.00226	0.99774	99,154	72	0.05270	0.94730	57,066
8	0.00249	0.99751	98,930	3	0.05752	0.94248	54,058
9	0.00274	0.99726	98,683	4	0.06279	0.93721	50,949
40	0.00301	0.99699	98,413	5	0.06853	0.93147	47,750
1	0.00330	0.99670	98,117	6	0.07480	0.92520	44,478
42	0.00363	0.99637	97,793	77	0.08164	0.91836	41,151
3	0.00398	0.99602	97,438	8	0.08910	0.91090	37,792
4	0.00437	0.99563	97,050	9	0.09725	0.90275	34,425
5	0.00479	0.99521	96,626	80	0.10613	0.89387	31,077
6	0.00525	0.99475	96,163	1	0.11582	0.88418	27,779
47	0.00575	0.99425	95,658	82	0.12641	0.87359	24,561
8	0.00629	0.99371	95,108	3	0.13795	0.86205	21,456
9	0.00687	0.99313	94,511	4	0.15055	0.84945	18,497
50	0.00753	0.99247	93,861	5	0.16429	0.83571	15,712
1	0.00824	0.99176	93,154	6	0.17929	0.82071	13,130
52	0.00901	0.99099	92,387	87	0.19563	0.80437	10,776
3	0.00985	0.99015	91,555	8	0.21348	0.78652	8,668
4	0.01077	0.98923	90,653	9	0.23296	0.76704	6,818
5	0.01178	0.98822	89,677	90	0.25421	0.74579	5,230
6	0.01287	0.98713	88,621	1	0.27741	0.72259	3,900
57	0.01407	0.98593	87,480	92	0.30272	0.69728	2,818
8	0.01537	0.98463	86,249	3	0.33033	0.66967	1,965
9	0.01680	0.98320	84,923	4	0.36046	0.63954	1,316
60	0.01835	0.98165	83,496	5	0.39333	0.60667	842
1	0.02004	0.97996	81,964	6	0.42920	0.57080	511
62	0.02189	0.97811	80,322	97	0.46834	0.53166	291
3	0.02391	0.97609	78,564	8	0.51105	0.48895	155
4	0.02611	0.97389	76,685	9	0.55765	0.44235	76
5	0.02851	0.97149	74,682	100	0.60849	0.39151	33
6	0.03113	0.96887	72,553	1	0.66397	0.33603	13
				102	0.72459	0.27541	4
				3	—	—	1

Table 6 contains, for every age between 32 and 102, the graduated values of q_x , the probability of an individual of age x dying within one year. In the next column are given the corresponding probabilities of survival p_x , or $1 - q_x$, i.e. the probability that a life age x will survive to age $x + 1$. Finally, table 6 contains the 'life table' proper. Starting with 100,000 'Fellows' of exact age 32, column (iii) of the table shows how many of these (l_x) will be expected to survive to each subsequent age x . (The recurrence relation is of course $l_{x+1} = p_x l_x$.) For instance, out of the original 100,000 entrants at age 32, rather more than 50,000 survive to age 74, and just under 25,000 live to attain the age of 82.

The expectation of life at age x is defined as the average number of years lived after age x ; that is, in the notation of the life table,

$$\frac{1}{l_x} \int_x^{\omega} l_t dt$$

(where ω denotes the oldest age of the table) or, to a sufficient approximation

$$\frac{1}{2} + \frac{1}{l_x} \sum_{x+1}^{\omega} l_t.$$

A few specimen values of the expectation, calculated from table 6, are quoted in table 7.

TABLE 7

age	expectation of
	life of Fellows of the Royal Society
	years
35	37.5
45	28.6
55	20.4
65	13.4
75	7.9
85	4.1
95	1.8

Our mortality investigation has covered a period of as long as 45 years, 1899–1943 inclusive. It is accordingly relevant to inquire whether any large secular changes were operative during this interval. While the small volume of data precludes detailed analysis, the total period was subdivided into two, 1 January 1899 to 31 December 1918, and 1 January 1919 to 31 December 1943, and the mortality experience in each was tested in the following manner. Suppose there were no secular trend in the mortality rate at a given age, say x . Then we should expect the ratio of deaths, occurring between ages x and $x + 1$, in the two subperiods, to be proportional to the 'years exposed to risk at age x ' in the respective periods. This hypothesis proved consistent with the facts for ages up to and including 74. Thereafter, it became rather discrepant. The results are summarized in table 8.

TABLE 8

ages	actual deaths		expected deaths if mortality rates were constant throughout the whole period	
	1899-1918	1919-1943	1899-1918	1919-1943
32-74	135	198	135	198
75-97	135	207	156	186
all ages	270	405	291	384

The inferences are that mortality rates at ages of 74 and less did not alter significantly throughout the years 1899-1944; but that the rates at ages of 75 and older may have become *heavier* in the latter part of the period. The χ^2 test indicates that the apparent increase in mortality at advanced ages was probably real. It may be mentioned that the change appeared to be most marked for the ages 75-80.

The conclusion that at some ages the vitality of Fellows is now less than in the first two decades of the century is at first sight surprising. It is not, however, an isolated phenomenon. The vitality of the country as a whole has, for instance, improved steadily over many years, *except* at the most advanced ages. Thus the Registrar-General (1936) showed that the male mortality rates at ages 70 and over were greater during 1921-30 than during the preceding decennium.

THE FUTURE FELLOWSHIP OF THE ROYAL SOCIETY

Assuming that the mortality of Fellows will in future be the same as that of 1899-1943, the life table of table 6 enables one to forecast the size and age distribution of the Society in the years ahead. This assumption may, in the light of the preceding paragraph, prove false, but it is considered that the available data are insufficient to justify a more elaborate hypothesis.

As regards Fellows who will enter the Society in future, it is supposed in the first instance that their ages will be distributed according to the final column of table 3. Four bases have been adopted—for purely illustrative purposes—viz. that the number elected annually will be

- (a) 20;
- (b) 25;
- (c) 30 for 5 years, reducing to 25 thereafter;
- (d) 35 for 5 years, reducing to 25 thereafter.

The results of the calculations are shown in figure 1. (For convenience, the calculations were made as at 1 January 1944.) A few comments may be offered in addition.

On each of the four bases, the total Fellowship will increase for some years, at a rate which is, at first, very sensitive to the number of annual elections (but is not at all sensitive to the ages of the new Fellows). Thereafter, the rate of growth will decline, and will eventually fall away to zero, as the age distribution of the Society becomes stable.

After 30–45 years (depending on the rate of election) the Society will reach a steady state, in which the number of deaths each year will, apart from random fluctuations, equal the number of annual elections, and in which the age constitution will remain constant.

If 35 or more candidates were admitted each year, for the next 5 years, the figure being reduced to 25 in later years, the total Fellowship would reach a maximum between 1975 and 1985, and would thereafter decline to the steady state population of about 650. On basis (d) this maximum would be about 660, and the slight decline from this figure to 650 might well be masked by random fluctuations.

The final steady state population is about 26 per annual election of 1 Fellow—assuming the age distribution at election of table 3.

It was pointed out to the author that, if the number of annual elections were raised, it might lead to some reduction in the average age of successful candidates for the Fellowship (despite the fact that the increase in 1938 of the number elected from 17 to 20 per year, has had no such effect yet—see table 3).

In order to illustrate the consequences of a reduction in the ages at election, it has been assumed that Fellows will henceforward be elected 3 years younger than was postulated before, i.e. the following hypothetical age distribution (table 9) of successful candidates has been adopted (cf. table 3):

TABLE 9

age at election	hypothetical proportion of Fellows elected
	%
27–36	20
37–46	40
47–56	30
57 and over	10
	<hr/> 100

The results of the calculation are shown in figure 2 for the two extreme cases considered, viz. that the number of new Fellows admitted will be

(a) 20 per year,

(d) 35 per year for the next 5 years, 25 per year thereafter.

Bases (b) and (c) would lead to the same final population—about 710 Fellows—as would (d). It will be observed that the suggested reduction in age would have a negligible effect on the rate of growth of the Society in the immediate future. It would, however, produce an increase in the final equilibrium population, equivalent to about $28\frac{1}{2}$ Fellows in the Society per annual election of 1 Fellow, compared with 26 Fellows per annual election of 1 Fellow, on the present age distribution of new entrants.

It has already been noted that the natural growth of the Society under present conditions will entail some further slight ageing of the Fellowship as a whole. The

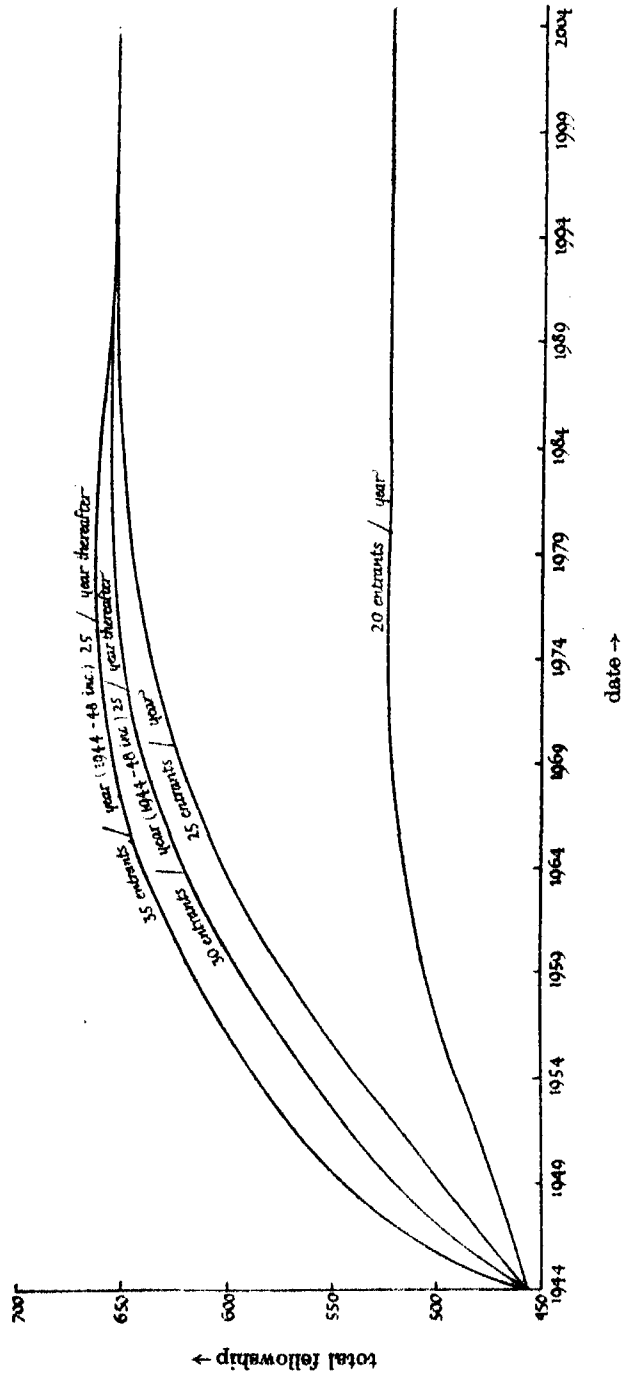


FIGURE 1. Estimated future Fellowship of the Royal Society. Assuming no change in age distribution of Fellows at election.

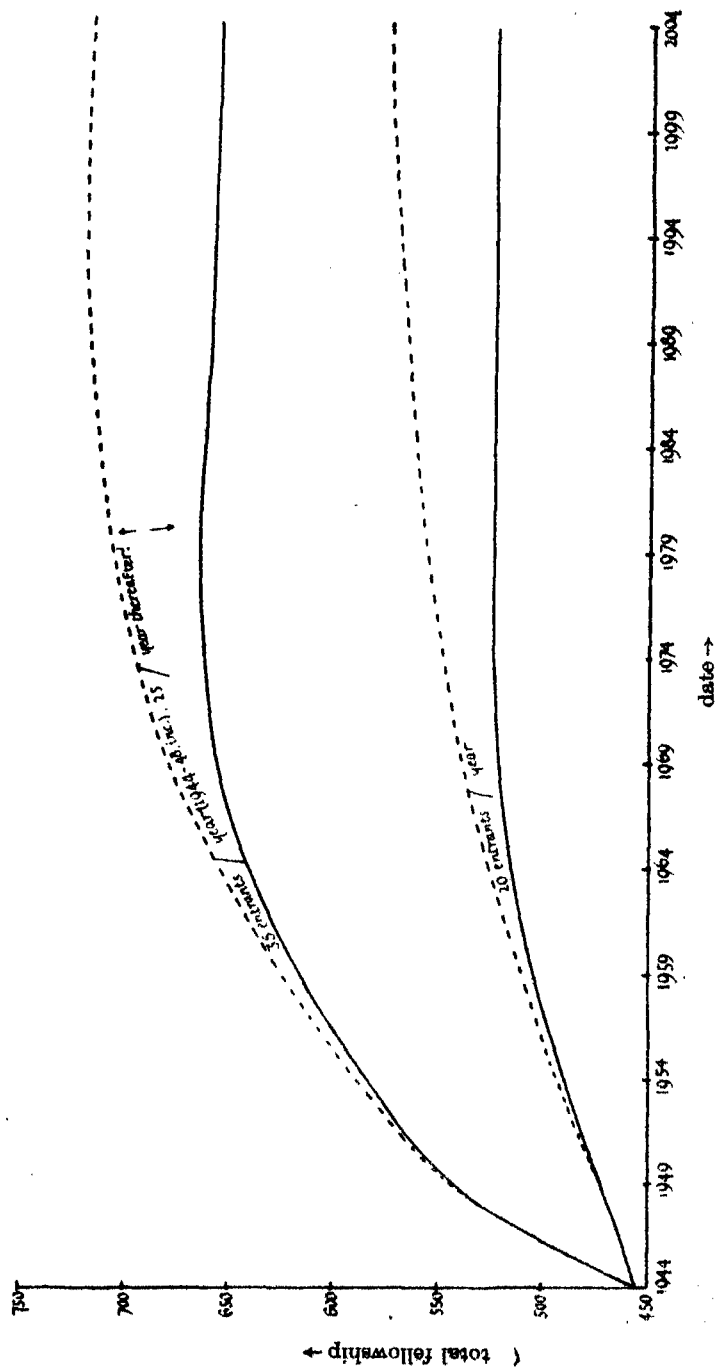


FIGURE 2. Estimated future Fellowship of the Royal Society. — assuming no change in age distribution at election.
 ... assuming a 3-year reduction in age of all Fellows elected in future.

extent of this movement, and the effect of electing younger candidates in future, are summarized in table 10.

TABLE 10. AGE DISTRIBUTION OF THE ROYAL SOCIETY

	proportion of Fellows of age		
	under 40	40-55	over 55
(1) actual distribution on 1 January 1944 %	3	32	65
(2) in steady state, assuming the present age distribution at election (table 3) %	3	27	70
(3) in steady state, assuming the modified age distribution at election (table 9) %	6	30	64

In conclusion it must be emphasized that the forecasts given above are necessarily tentative. The mortality which the Royal Society will experience during the remainder of this century may differ sufficiently from that of the past to belie the estimates made. Moreover, no consideration has been given to such factors as the possible election of women Fellows, whose expectation of life and whose ages at election may be very different from those of their male colleagues; or the possibly greater representation of Dominions, Indian and Colonial scientists, whose mortality experience is unlikely to conform to the standard prevailing in this country. Their effects, while at the moment unpredictable, may well be far-reaching on the actuarial constitution of the Royal Society.

This investigation was undertaken on the suggestion of the Council of the Royal Society. The author wishes to thank Mr Griffith Davies, the Assistant Secretary of the Royal Society, and his staff for their help in extracting and tabulating the data. Thanks are also due to Professor A. V. Hill, Professor P. M. S. Blackett, and Dr E. C. Bullard, for their advice and helpful criticism, and to Miss J. M. Westcombe for checking the computations.

REFERENCES

- Elderton, W. P. & Fippard, R. C. 1935 *The construction of mortality and sickness tables*. London: A. and C. Black, Ltd.
- Elderton, W. P. & Oakley, H. J. P. 1924 *The mortality of annuitants, 1900-1920*. London: C. and E. Layton.
- Gompertz, B. 1825 *Phil. Trans.* 115, 513.
- Hardy, G. F. 1909 *The theory of the construction of tables of mortality*. Cambridge University Press.
- Hill, A. V. 1939 *Notes and records of the Royal Society*, p. 71.
- Makeham, W. H. 1860 *J. Inst. Actu.* 8, 301.
- Registrar-General 1936 *Decennial Supplement, England and Wales, 1931*. London: H.M. Stationery Office.
- Schuster, A. 1922 *Proc. Roy. Soc. A*, 107, 368.
- Seal, H. L. 1941 *J. Inst. Actu.* 71, 5.
- Strachey, R. 1892 *Proc. Roy. Soc. A*, 51, 463.
- 1935 *Mortality of assured lives, 1924-29*. (Extracts and Discussions.) Cambridge University Press.

INDEX TO VOLUME 184 (A)

- Age constitution and the future fellowship of the Royal Society (Solomon), 464.
- Aliphatic hydrocarbons, catalytic cyclisation, I (Herington & Rideal), 434.
- Aliphatic hydrocarbons, catalytic cyclisation, II (Herington & Rideal), 447.
- Anderson, J. S. & Morton, M. C. The electrical conductivity of stannous sulphide, 83.
- Auger, P. Some new investigations on cosmic ray showers (Abstract), 2.
- Bakerian Lecture. On relaxation methods: a mathematics for engineering science (Southwell), 253.
- Carlisle, C. H. & Crowfoot, D. The crystal structure of cholesteryl iodide, 64.
- Complex potentials in two-dimensional elasticity, I (Stevenson), 129.
- Complex potentials in two-dimensional elasticity, II (Stevenson), 218.
- Cosmic ray showers, some new investigations (Auger), 2.
- Crowfoot, D. *See* Carlisle & Crowfoot.
- Crystal structure and cholesteryl iodide (Carlisle & Crowfoot), 64.
- Eckersley, T. L. & Farmer, F. T. Short period fluctuations in the characteristics of wireless echoes from the ionosphere, 196.
- Electrical conductivity of stannous sulphide (Anderson & Morton), 83.
- Elkes, J. J., Frazer, A. C., Schulman, J. H. & Stewart, H. C. Reversible adsorption of proteins at the oil/water interface. Preferential adsorption of proteins at charged oil/water interfaces, 102.
- Farmer, F. T. *See* Eckersley & Farmer.
- Frazer, A. C. *See* Elkes & others.
- Garlick, J. F. G. & Wilkins, M. H. F. Short period phosphorescence and electron traps, 408.
- Green, A. E. Stress systems in isotropic and aeolotropic plates, V, 231.
- Green, A. E. Stress systems in aeolotropic plates, VI, 289.
- Green, A. E. Stress systems in aeolotropic plates, VII, 301.
- Green, A. E. & Taylor, G. I. Stress systems in aeolotropic plates, III, 181.
- Herington, E. F. G. & Rideal, E. K. On the catalytic cyclisation of aliphatic hydrocarbons, I, 434.
- Herington, E. F. G. & Rideal, E. K. On the catalytic cyclisation of aliphatic hydrocarbons, II, 447.
- Infra-red spectra of polymers and related monomers, I (Thompson & Torkington), 3.
- Infra-red spectra of polymers and related monomers, II (Thompson & Torkington), 21.
- Joliot, F. Métabolisme de la thyroxine et de l'iode ionique (Abstract), 1.
- Joliot-Curie, Irène. Questions of measurements and standardization in radioactivity (Abstract), 2.
- Métabolisme de la thyroxine et de l'iode ionique (Joliot), 1.
- Mokhtar, K. & Richardson, E. G. Supersonic dispersion in gases. II. Air containing water vapour, 117.
- Morton, M. C. *See* Anderson & Morton.

- Phosphorescence and electron traps (Randall & Wilkins), 365.
 Phosphorescence of various solids (Randall & Wilkins), 347.
 Proteins, reversible adsorption at oil/water interface (Elkes & others), 102.
- Radioactivity. Questions of measurements and standardization (Joliot-Curie), 2.
 Randall, J. T. & Wilkins, M. H. F. Phosphorescence and electron traps, 365.
 Randall, J. T. & Wilkins, M. H. F. The phosphorescence of various solids, 347.
 Relaxation methods: a mathematics for engineering science. Bakerian Lecture (Southwell), 253.
 Reversible adsorption of proteins at the oil/water interface. Preferential adsorption of proteins at charged oil/water interfaces (Elkes & others), 102.
 Richardson, E. G. *See* Mokhtar & Richardson.
 Rideal, E. K. *See* Herington & Rideal.
- Schulman, H. J. *See* Elkes & others.
 Short period fluctuations in the characteristics of wireless echoes from the ionosphere (Eckersley & Farmer), 196.
 Short period phosphorescence and electron traps (Garlick & Wilkins), 408.
 Solomon, I. The age constitution and the future fellowship of the Royal Society, 464.
 Southwell, R. V. Bakerian Lecture. On relaxation methods: a mathematics for engineering science, 253.
 Stevenson, A. C. Complex potentials in two-dimensional elasticity, I, 129.
 Stevenson, A. C. Complex potentials in two-dimensional elasticity, II, 218.
 Stewart, H. C. *See* Elkes & others.
- Stress systems in aeolotropic plates, III (Green & Taylor), 181.
 Stress systems in isotropic and aeolotropic plates, V (Green), 231.
 Stress systems in aeolotropic plates, VI (Green), 289.
 Stress systems in aeolotropic plates, VII (Green), 301.
 Supersonic dispersion in gases, II. Air containing water vapour (Mokhtar & Richardson), 117.
- Taylor, G. I. *See* Green & Taylor.
 Thompson, H. W. & Torkington, P. The infra-red spectra of polymers and related monomers, I, 3.
 Thompson, H. W. & Torkington, P. The infra-red spectra of polymers and related monomers, II, 21.
 Tolansky, S. The topography of crystal faces. I. The topography of a (100) face of a left-handed quartz crystal, 41.
 Tolansky, S. The topography of crystal faces. II. The topography of cleavage faces of mica and selenite, 51.
 Topography of crystal faces. I. The topography of a (100) face of a left-handed quartz crystal (Tolansky), 41.
 Topography of crystal faces. II. The topography of cleavage faces of mica and selenite (Tolansky), 51.
 Torkington, P. *See* Thompson & Torkington.
- Wilkins, M. H. F. *See* Garlick & Wilkins.
 Wilkins, M. H. F. *See* Randall & Wilkins.

INDIAN AGRICULTURAL RESEARCH
INSTITUTE LIBRARY
NEW DELHI.

[illegible]

S. C. P.—1/8/47 P. J.-3-5-48-2000